

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 15, 2001, 07:55:57 ; Search time 98.8 Seconds

(without alignments)
114.395 Million cell updates/sec

Title: US-09-663-020-8

Perfect score: 18

Sequence: 1 ggcggagcttcagctgag 18

Scoring table:

OLIGO_NUC

Searched: 730101 seqs, 313950809 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

N.Geneseq.0601.*
1: /SIDSL/gcgdata/geneseq/geneseq/NA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseq/NA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseq/NA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseq/NA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseq/NA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseq/NA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseq/NA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseq/NA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseq/NA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseq/NA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseq/NA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseq/NA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseq/NA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseq/NA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseq/NA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseq/NA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseq/NA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseq/NA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT:*
22: /SIDSL/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	18	100.0	82	21	AAC23183
c 2	18	100.0	113	21	AAC23260
c 3	18	100.0	126	21	AAC18414
c 4	18	100.0	131	21	AAC28492
c 5	18	100.0	134	21	AAC04894
c 6	18	100.0	135	21	AAC18209
c 7	18	100.0	138	21	AAC11987
c 8	18	100.0	138	21	AAC15609
c 9	18	100.0	151	21	AAC15690
c 10	18	100.0	167	21	AAC05392
c 11	18	100.0	172	21	AAC11951

c 12	18	100.0	182	21	AAC04655	Human secreted pro
c 13	18	100.0	185	21	AAC16191	Human secreted pro
c 14	18	100.0	186	21	AAC22610	Human secreted pro
c 15	18	100.0	187	21	AAC22471	Human secreted pro
c 16	18	100.0	188	21	AAC21755	Human secreted pro
c 17	18	100.0	191	21	AAC15430	Human secreted pro
c 18	18	100.0	207	21	AAC15123	Human secreted pro
c 19	18	100.0	208	21	AAC15912	Human secreted pro
c 20	18	100.0	212	21	AAC12942	Human secreted pro
c 21	18	100.0	212	21	AAC12942	Human secreted pro
c 22	18	100.0	212	21	AAC22434	Human secreted pro
c 23	18	100.0	227	21	AAC24382	Human secreted pro
c 24	18	100.0	231	21	AAC27982	Human secreted pro
c 25	18	100.0	232	21	AAC00600	Human secreted pro
c 26	18	100.0	232	21	AAC31476	Human secreted pro
c 27	18	100.0	242	21	AAC27480	Human secreted pro
c 28	18	100.0	244	21	AAC28162	Human secreted pro
c 29	18	100.0	247	21	AAC02486	Human secreted pro
c 30	18	100.0	248	21	AAC10373	Human secreted pro
c 31	18	100.0	251	21	AAC15555	Human secreted pro
c 32	18	100.0	256	20	AAV89497	Human secreted pro
c 33	18	100.0	266	21	AAA00011	Human secreted pro
c 34	18	100.0	271	21	AAC14406	Human secreted pro
c 35	18	100.0	277	21	AAC05484	Human secreted pro
c 36	18	100.0	289	20	AAV87246	Human secreted pro
c 37	18	100.0	300	20	AAZ13338	Human secreted pro
c 38	18	100.0	300	20	AAZ13338	Human secreted pro
c 39	18	100.0	310	21	AAC04771	Human secreted pro
c 40	18	100.0	316	21	AAC06081	Human secreted pro
c 41	18	100.0	320	18	AAT94758	Human secreted pro
c 42	18	100.0	320	18	AAT84749	Human secreted pro
c 43	18	100.0	320	21	AAC05638	Human secreted pro
c 44	18	100.0	324	21	AAC26578	Human secreted pro
c 45	18	100.0	334	21	AAA43586	Human secreted exp

ALIGNMENTS

RESULT 1
AAC23183/C
AAC23183 standard; CDNA; 82 BP.
AC AAC23183;
XX
DT 06-OCT-2000 (first entry)
XX
XX Human secreted protein 5' EST, SEQ ID NO: 27258.
DE Human: 5' EST; expressed sequence tag; secreted protein; CDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
DR
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 27258; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.

SO Sequence 82 BP; 14 A; 33 C; 19 G; 16 T; 0 other;

Query Match 100.0%; Score 18; DB 21; Length 82;
 Best Local Similarity 100.0%; Pred. No. 0.15; Mismatches 0; Gaps 0;
 Matches 18; Conservative 0; Indels 0;

OY 1 ggcggagcttcagtgag 18
 |||||
 DB 22 GCGGAGCTTCAGTGAG 5

RESULT 2
 AAC23260/c
 ID AAC23260 standard; cDNA; 113 BP.

AC AAC23260;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 27335.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KM gene therapy; chromosome mapping; ss.

OS Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

PI Dumas Milne Edwards J, Duclet A, Giordano J;

DR MPI; 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 PS Claim 1: SEQ ID 27335; 71bp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be

CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.

SO Sequence 113 BP; 19 A; 42 C; 22 G; 30 T; 0 other;

Query Match 100.0%; Score 18; DB 21; Length 113;
 Best Local Similarity 100.0%; Pred. No. 0.15; Mismatches 0; Gaps 0;
 Matches 18; Conservative 0; Indels 0;

OY 1 ggcggagcttcagtgag 18
 |||||
 DB 22 GCGGAGCTTCAGTGAG 5

RESULT 3

AAC18414/c
 ID AAC18414 standard; cDNA; 126 BP.

AC AAC18414;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 22489.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KM gene therapy; chromosome mapping; ss.

OS Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

PI Dumas Milne Edwards J, Duclet A, Giordano J;

DR MPI; 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 PS Claim 1: SEQ ID 22489; 71bp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.

SO Sequence 126 BP; 23 A; 49 C; 30 G; 24 T; 0 other;

Query Match 100.0%; Score 18; DB 21; Length 126;
 Best Local Similarity 100.0%; Pred. No. 0.15; Mismatches 0; Gaps 0;
 Matches 18; Conservative 0; Indels 0;

OY 1 ggcggagcttcagtgag 18
 |||||||
 DB 50 GCGGAGCTTCAGTGAG 33

RESULT 4
 AAC28492/c
 ID AAC28492 standard; cDNA; 131 BP.
 XX

AC AAC28492;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 32567.

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KM gene therapy; chromosome mapping; ss.

OS Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 990S-0122487.

PA (GEST) GENSET.

PI Dumas Mline Edwards J, Duclert A, Giordano J;

DR WPI: 2000-500381/45.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 diagnostic, forensic, gene therapy and chromosome mapping procedures -
 PS Claim 1; SEQ ID 32567; 71pp + CD-ROM; English.

CC The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.

SO Sequence 131 BP; 22 A; 46 C; 35 G; 28 T; 0 other;

Query Match 100.0%; Score 18; DB 21; Length 131;
 Best Local Similarity 100.0%; Pred. No. 0.15;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggcggagcttcagtgag 18
 |||||||
 DB 56 GCGGAGCTTCAGTGAG 39

RESULT 5
 AAC04894/c
 ID AAC04894 standard; cDNA; 134 BP.
 XX

AC AAC04894;

XX 06-OCT-2000 (first entry)
 DT Human secreted protein 5' EST, SEQ ID NO: 8969.
 DE Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KM gene therapy; chromosome mapping; ss.

OS Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 990S-0122487.

PA (GEST) GENSET.

PI Dumas Mline Edwards J, Duclert A, Giordano J;

DR WPI: 2000-500381/45.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 diagnostic, forensic, gene therapy and chromosome mapping procedures -
 PS Claim 1; SEQ ID 8969; 71pp + CD-ROM; English.

CC The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.

SO Sequence 134 BP; 18 A; 49 C; 35 G; 28 T; 4 other;

Query Match 100.0%; Score 18; DB 21; Length 134;
 Best Local Similarity 100.0%; Pred. No. 0.15;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggcggagcttcagtgag 18
 |||||||
 DB 60 GCGGAGCTTCAGTGAG 43

RESULT 6
 AAC18209/c
 ID AAC18209 standard; cDNA; 135 BP.
 XX

AC AAC18209;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 22284.

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KM gene therapy; chromosome mapping; ss.

OS Homo sapiens.

PN EP1033401-A2.


```

XX 26-FEB-1999; 99US-0122487.
XX (GEST ) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI: 2000-500381/45.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 1; SEQ ID 20266; 71bp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX identified within the present sequence. The 5' ESTs were prepared from
XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX sequences usually correspond mainly to the 3' untranslated region (UTR)
XX of the mRNA because they are often obtained from oligo-dT primed cDNA
XX libraries. Such ESTs are not well suited for isolating cDNA sequences
XX derived from the 5' ends of mRNAs and even in those cases where longer
XX cDNA sequences have been obtained, the full 5' UTR is rarely included.
XX 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
XX used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
XX in diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX They are used to obtain upstream regulatory sequences and to design
XX expression and secretion vectors.
XX
XX Sequence 185 BP; 56 A; 44 C; 54 G; 31 T; 0 other;
XX
XX Query Match 100.0%; Score 18; DB 21; Length 185;
XX Best Local Similarity 100.0%; Pred. No. 0.14;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 ggcggagcttcagtgag 18
XX ||||||||||||||||
XX 102 ggcggagcttcagtgag 119
XX
XX RESULT 14
XX AAC22610/C
XX ID AAC22610 standard; cDNA; 186 BP.
XX
XX AAC22610;
XX
XX 06-OCT-2000 (first entry)
XX
XX Human secreted protein 5' EST, SEQ ID NO: 26685.
XX
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX gene therapy; chromosome mapping; ss.
XX
XX Homo sapiens.
XX
XX EP1033401-A2.
XX
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-0200610.
XX
XX 26-FEB-1999; 99US-0122487.
XX
XX (GEST ) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI: 2000-500381/45.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for

```

```

PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 1; SEQ ID 26685; 71bp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX identified within the present sequence. The 5' ESTs were prepared from
XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX sequences usually correspond mainly to the 3' untranslated region (UTR)
XX of the mRNA because they are often obtained from oligo-dT primed cDNA
XX libraries. Such ESTs are not well suited for isolating cDNA sequences
XX derived from the 5' ends of mRNAs and even in those cases where longer
XX cDNA sequences have been obtained, the full 5' UTR is rarely included.
XX 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
XX used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
XX in diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX They are used to obtain upstream regulatory sequences and to design
XX expression and secretion vectors.
XX
XX Sequence 186 BP; 34 A; 47 C; 40 G; 65 T; 0 other;
XX
XX Query Match 100.0%; Score 18; DB 21; Length 186;
XX Best Local Similarity 100.0%; Pred. No. 0.14;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 ggcggagcttcagtgag 18
XX ||||||||||||||||
XX 118 ggcggagcttcagtgag 101
XX
XX RESULT 15
XX AAC22471/C
XX ID AAC22471 standard; cDNA; 187 BP.
XX
XX AAC22471;
XX
XX 06-OCT-2000 (first entry)
XX
XX Human secreted protein 5' EST, SEQ ID NO: 26546.
XX
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX gene therapy; chromosome mapping; ss.
XX
XX Homo sapiens.
XX
XX EP1033401-A2.
XX
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-0200610.
XX
XX 26-FEB-1999; 99US-0122487.
XX
XX (GEST ) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI: 2000-500381/45.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 1; SEQ ID 26546; 71bp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX identified within the present sequence. The 5' ESTs were prepared from
XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX sequences usually correspond mainly to the 3' untranslated region (UTR)
XX of the mRNA because they are often obtained from oligo-dT primed cDNA
XX libraries. Such ESTs are not well suited for isolating cDNA sequences

```

CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.

XX
 SQ Sequence 187 BP; 34 A; 39 C; 39 G; 75 T; 0 other;

Query Match

Best Local Similarity 100.0%; Score 18; DB 21; Length 187;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggcgagccttcagtgag 18
 |||
 Db 183 ggcgagccttcagtgag 166

Search completed: November 15, 2001, 07:55:58
 Job time: 7258 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 15, 2001, 05:55:00 ; Search time 98.8 Seconds

(Without alignments)
152.527 Million cell updates/sec

Title: US-09-663-020-7

Sequence: 1 tcagattatttgggtcaccac 24

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 730101 seqs, 313950809 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_0601.*

```
1: /SIDSL/gcgdata/geneseq/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/NA2001.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16	66.7	774	20	AA61572
2	16	66.7	933	20	AA61571
3	16	66.7	910715	20	AA20248
4	15	62.5	743	21	AA30331
5	15	62.5	1310	21	AA50076
6	15	62.5	1313	21	AA37786
7	15	62.5	1392	20	AA91624
8	15	62.5	1836	19	AAV33461
9	15	62.5	2083	22	AA93852
10	15	62.5	2385	21	AA39981
11	15	62.5	2403	21	AA39953

12	15	62.5	2418	21	AAA3979
13	15	62.5	2490	21	AAA3975
14	15	62.5	2523	21	AAA3978
15	15	62.5	2562	21	AAA3980
16	15	62.5	2593	21	AAA3976
17	15	62.5	2700	21	AAA3977
18	14	58.3	271	21	AA06457
19	14	58.3	271	21	AA15964
20	14	58.3	367	21	AA05418
21	14	58.3	459	21	AA05274
22	14	58.3	477	21	AA03740
23	14	58.3	486	17	AA027168
24	14	58.3	486	17	AA027168
25	14	58.3	799	22	AA022424
26	14	58.3	874	19	AA037365
27	14	58.3	896	20	AA037677
28	14	58.3	927	22	AA022520
29	14	58.3	965	22	AA022423
30	14	58.3	1077	20	AA098000
31	14	58.3	1100	22	AA023651
32	14	58.3	1156	22	AA033243
33	14	58.3	1300	21	AA061504
34	14	58.3	1502	17	AA027165
35	14	58.3	1502	17	AA027165
36	14	58.3	1572	22	AA092064
37	14	58.3	1647	22	AA022517
38	14	58.3	1679	19	AA033885
39	14	58.3	1747	15	AA062810
40	14	58.3	1781	21	AA069636
41	14	58.3	2045	20	AA019033
42	14	58.3	2537	16	AA092571
43	14	58.3	2539	21	AA088555
44	14	58.3	3309	21	AA040073
45	14	58.3	3546	20	AA042844

ALIGNMENTS

```
RESULT 1
AA61572
ID AAX61572 standard; DNA: 774 BP.
AC AAX61572;
XX
DT 19-JUL-1999 (first entry)
XX
DE B. burgdorferi antigenic protein coding sequence, t617.nt.
XX
KW Antigenic protein; vaccine; Lyme disease; Infection; detection; ss.
XX
OS Borrelia burgdorferi.
XX
PN WO9859071-A1.
XX
PD 30-DEC-1998.
XX
PF 18-JUN-1998; 98WO-0512718.
XX
PR 03-SEP-1997; 97US-0057483.
PR 20-JUN-1997; 97US-0050359.
PR 22-JUL-1997; 97US-0053344.
PR 22-JUL-1997; 97US-0053377.
XX
PA (HDM-) HUMAN GENOME SCI INC.
XX (MEDI-) MEDIMUNE INC.
XX
PI Choi GH, Ertvin AL, Hanson MS, Lathigra R;
XX WPI: 1999-189980/16.
XX P-PSDB: AAY19875.
XX
PT New isolated Borrelia burgdorferi nucleic acids - used to develop
```

PT products for the diagnosis, prevention and treatment of diseases
 caused by *Borrelia*, particularly Lyme disease

PS Claim 1: Page 100; 275pp; English.

CC This sequence encodes a *Borrelia burgdorferi* (Bb) protein of the
 CC invention, which is suitable for use in a vaccine. The Bb polypeptides
 CC can be used in vaccines for eliciting protective antibodies to members of
 CC the *Borrelia* genus, particularly for the use against Lyme disease in
 CC humans and animals. They can be used for preventing or attenuating an
 CC infection caused by a member of the *Borrelia* genus. The products can also
 CC be used for detection of members of the *Borrelia* genus.

XX Sequence 774 BP; 218 A; 74 C; 124 G; 358 T; 0 other;

Query Match 66.7%; Score 16; DB 20; Length 774;
 Best Local Similarity 100.0%; Pred. No. 4.1;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tcagattatttggg 16
 |||||
 Db 36 tcagattatttggg 51

RESULT 2

AAK61571
 ID AAK61571 standard; DNA; 933 BP.

XX AAK61571;

DT 19-JUL-1999 (first entry)

DE B. burgdorferi antigenic protein coding sequence, f617.nt.

KW Antigenic protein; vaccine; Lyme disease; infection; detection; ss.

OS *Borrelia burgdorferi*.

XX WO9859071-A1.

PN 30-DEC-1998.

PF 18-JUN-1998; 98WO-US12718.

PR 03-SEP-1997; 97US-0057483.

PR 20-JUN-1997; 97US-0050359.

PR 22-JUL-1997; 97US-0053344.

XX 22-JUL-1997; 97US-0053377.

PA (HUMA-) HUMAN GENOME SCI INC.

XX (MEDI-) MEDIMUNE INC.

XX ChOI GH, Erwin AL, Hanson MS, Lathigra R;

DR WPI: 1999-189980/16.

DR P-PSDB: AAY19874.

PT New isolated *Borrelia burgdorferi* nucleic acids - used to develop

PT products for the diagnosis, prevention and treatment of diseases

PT caused by *Borrelia*, particularly Lyme disease

XX Claim 1: Page 100; 275pp; English.

CC This sequence encodes a *Borrelia burgdorferi* (Bb) protein of the
 CC invention, which is suitable for use in a vaccine. The Bb polypeptides
 CC can be used in vaccines for eliciting protective antibodies to members of
 CC the *Borrelia* genus, particularly for the use against Lyme disease in
 CC humans and animals. They can be used for preventing or attenuating an
 CC infection caused by a member of the *Borrelia* genus. The products can also
 CC be used for detection of members of the *Borrelia* genus.

XX Sequence 933 BP; 256 A; 90 C; 146 G; 441 T; 0 other;

Query Match 66.7%; Score 16; DB 20; Length 933;
 Best Local Similarity 100.0%; Pred. No. 4.1;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tcagattatttggg 16
 |||||
 Db 195 tcagattatttggg 210

RESULT 3

AAK20248
 ID AAK20248 standard; DNA; 910715 BP.

XX AAK20248;

DT 04-MAY-1999 (first entry)

DE *Borrelia burgdorferi* polynucleotide sequence #1.

XX *Borrelia burgdorferi*; spirochete; bacterium; pathogen; Lyme disease;

KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;

XX infection; diagnosis; characterisation; detection; ds.

OS *Borrelia burgdorferi*.

XX WO9858943-A1.

PN 30-DEC-1998.

PF 18-JUN-1998; 98WO-US12764.

PR 03-SEP-1997; 97US-0057483.

PR 20-JUN-1997; 97US-0050359.

PR 22-JUL-1997; 97US-0053344.

XX 22-JUL-1997; 97US-0053377.

PA (HUMA-) HUMAN GENOME SCI INC.

XX (MEDI-) MEDIMUNE INC.

XX Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO;

PI White OR;

DR WPI: 1999-081217/07.

PT New isolated *Borrelia burgdorferi* nucleic acids - used to develop

PT products for the detection, diagnosis, characterisation, prevention

PT and therapy of infections, particularly Lyme disease

XX Claim 1: Page 157-671; 1128pp; English.

XX AAK20248 to AAK20402 represent polynucleotide sequences isolated from

CC *Borrelia burgdorferi* (Bb). Products derived from Bb can be used for

CC the detection, diagnosis, characterisation, prevention and therapy of

CC Bb infections; e.g. Lyme disease. They can also be used for the

CC production of biosynthetic products, e.g. enzymes. *Borrelia* belongs

CC to a family of motile, spiral-shaped bacteria called Spirochetes.

CC Spirochetes are pathogenic in humans and *Borrelia* causes epidemic and

CC endemic relapsing fever, and Lyme borreliosis, more commonly known as

XX Lyme disease.

XX Sequence 910715 BP; 327171 A; 129646 C; 130753 G; 323091 T; 54 other;

Query Match 66.7%; Score 16; DB 20; Length 910715;
 Best Local Similarity 100.0%; Pred. No. 5.2;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tcagattatttggg 16
 |||||
 Db 587127 tcagattatttggg 587142

```
RESULT 4
AAA0331/C
ID AAA0331 standard; cDNA; 743 BP.
XX
AC AAA0331;
XX
DT 05-SEP-2000 (first entry)
XX
DE Human thymic stromal lymphopoietin (TSLP) coding sequence.
XX
KW Human; thymic stromal lymphopoietin; TSLP; B cell maturation;
KW immune regulation; cell proliferation; cell differentiation; cell death;
KW cell migration; cell-to-cell interaction; inflammatory response;
KW chromosome 5q21-22; Gardner syndrome; adenomatous polyposis coli;
KW hereditary desmoid disease; Turcot syndrome; colorectal cancer; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 200..679
FT /tag="a
FT /product="TSLP"
XX
PN WO200029581-A1.
XX
PD 25-MAY-2000.
XX
PF 12-NOV-1999; 99WO-US27069.
XX
PR 13-NOV-1998; 98US-0108452.
XX
PA (IMMUNEX ) IMMUNEX CORP.
XX
PI Slims J, Lyman S, McKenna H, Armstrong A;
XX
DR WPI; 2000-387794/33.
XX
DR P-PSDB; AAY97365.
XX
PT New human thymic stromal lymphopoietin (TSLP) polypeptide useful for
PT stimulating lymphocyte development and proliferation
XX
PS Claim 2; Fig 1; 78pp; English.
XX
CC The present sequence is the coding sequence for human thymic
CC stromal lymphopoietin (TSLP). It was isolated by searching an EST
CC library for sequences similar to the murine TSLP sequence. The protein
CC is involved in the growth and differentiation of B and T cells. It can
CC be used to study processes such as immune regulation, cell
CC proliferation, cell death, cell migration, cell-to-cell interaction and
CC inflammatory responses. The nucleic acid can be used to identify human
CC chromosome 5, to map genes along this chromosome, to identify genes on
CC this chromosome which are associated with diseases, including Gardner
CC syndrome, adenomatous polyposis coli, hereditary desmoid disease, Turcot
CC syndrome and colorectal cancer, and to inhibit or induce B and T cell
CC proliferation.
XX
SQ Sequence 743 BP; 226 A; 166 C; 161 G; 190 T; 0 other;
```

```
Query Match 63.5%; Score 15; DB 21; Length 743;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 4 gattatttggct 18
|||||
Db 721 GATTATTTCGGCT 707
```

```
RESULT 5
AAC50076
ID AAC50076 standard; DNA; 1310 BP.
XX
```

```
AC AAC50076;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 63506.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0130891.
XX 30-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 04-MAY-1999; 99US-0132407.
XX 05-MAY-1999; 99US-0132484.
XX 06-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 07-MAY-1999; 99US-0132487.
XX 11-MAY-1999; 99US-0132863.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 18-MAY-1999; 99US-0134370.
XX 19-MAY-1999; 99US-0134768.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.
```

PR	26-AUG-1999;	9905-0150884.
PR	27-AUG-1999;	9905-0151065.
PR	27-AUG-1999;	9905-0151066.
PR	27-AUG-1999;	9905-0151080.
PR	30-AUG-1999;	9905-0151303.
PR	31-AUG-1999;	9905-0151438.
PR	01-SEP-1999;	9905-0151930.
PR	07-SEP-1999;	9905-0152363.
PR	10-SEP-1999;	9905-0153070.
PR	13-SEP-1999;	9905-0153758.
PR	15-SEP-1999;	9905-0154018.
PR	16-SEP-1999;	9905-0154039.
PR	20-SEP-1999;	9905-0154779.
PR	22-SEP-1999;	9905-0155139.
PR	23-SEP-1999;	9905-0155486.
PR	24-SEP-1999;	9905-0155659.
PR	28-SEP-1999;	9905-0156458.
PR	29-SEP-1999;	9905-0156596.
PR	04-OCT-1999;	9905-0157117.
PR	05-OCT-1999;	9905-0157753.
PR	06-OCT-1999;	9905-0157865.
PR	07-OCT-1999;	9905-0158029.
PR	08-OCT-1999;	9905-0158232.
PR	12-OCT-1999;	9905-0158369.
PR	13-OCT-1999;	9905-0159293.
PR	13-OCT-1999;	9905-0159294.
PR	13-OCT-1999;	9905-0159295.
PR	14-OCT-1999;	9905-0159329.
PR	14-OCT-1999;	9905-0159330.
PR	14-OCT-1999;	9905-0159331.
PR	14-OCT-1999;	9905-0159637.
PR	14-OCT-1999;	9905-0159638.
PR	18-OCT-1999;	9905-0159584.
PR	21-OCT-1999;	9905-0160741.
PR	21-OCT-1999;	9905-0160767.
PR	21-OCT-1999;	9905-0160767.
PR	21-OCT-1999;	9905-0160768.
PR	21-OCT-1999;	9905-0160770.
PR	21-OCT-1999;	9905-0160814.
PR	21-OCT-1999;	9905-0160815.
PR	22-OCT-1999;	9905-0160980.
PR	22-OCT-1999;	9905-0160981.
PR	22-OCT-1999;	9905-0160989.
PR	25-OCT-1999;	9905-0161405.
PR	25-OCT-1999;	9905-0161405.
PR	25-OCT-1999;	9905-0161406.
PR	26-OCT-1999;	9905-0161359.
PR	26-OCT-1999;	9905-0161360.
PR	26-OCT-1999;	9905-0161361.
PR	28-OCT-1999;	9905-0161920.
PR	28-OCT-1999;	9905-0161992.
PR	28-OCT-1999;	9905-0161993.
PR	29-OCT-1999;	9905-0162142.

```

Query Match          62.5%; Score 15; DB 21; Length 1310;
Best Local Similarity 100.0%; Pried.No. 15;
Matches    15; Conservative    0; Mismatches    0; Indels    0; Gaps    0;

Qy      5 attcatttggcgctt 19
         |||
Db       1290 attcatttggcgctt 1304

RESULT 6
AAC37786 AAC37786 standard; DNA; 1313 BP.
XX
XX
AC
AC AAC37786;
XX
DF 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 18660.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
```


KW protein identification; signal transduction pathway;
KM metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.
OS
XX EPI033405-A2.
PN
XX
PD 06-SEP-2000.
PF
XX 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 990S-0121825.
PR 05-MAR-1999; 990S-0123180.
PR 09-MAR-1999; 990S-0123548.
PR 23-MAR-1999; 990S-0125788.
PR 25-MAR-1999; 990S-0126264.
PR 29-MAR-1999; 990S-0126785.
PR 01-APR-1999; 990S-0127462.
PR 06-APR-1999; 990S-0128234.
PR 08-APR-1999; 990S-0128714.
PR 16-APR-1999; 990S-0128845.
PR 19-APR-1999; 990S-0130077.
PR 21-APR-1999; 990S-0130449.
PR 23-APR-1999; 990S-0130510.
PR 28-APR-1999; 990S-0130891.
PR 30-APR-1999; 990S-0131449.
PR 30-APR-1999; 990S-0132048.
PR 04-MAY-1999; 990S-0132407.
PR 05-MAY-1999; 990S-0132484.
PR 06-MAY-1999; 990S-0132485.
PR 06-MAY-1999; 990S-0132486.
PR 07-MAY-1999; 990S-0132863.
PR 11-MAY-1999; 990S-0134256.
PR 14-MAY-1999; 990S-0134218.
PR 14-MAY-1999; 990S-0134219.
PR 14-MAY-1999; 990S-0134221.
PR 18-MAY-1999; 990S-0134370.
PR 19-MAY-1999; 990S-0134768.
PR 19-MAY-1999; 990S-0134941.
PR 20-MAY-1999; 990S-0135124.
PR 21-MAY-1999; 990S-0135353.
PR 24-MAY-1999; 990S-0135629.
PR 25-MAY-1999; 990S-0136021.
PR 27-MAY-1999; 990S-0136392.
PR 28-MAY-1999; 990S-0136782.
PR 01-JUN-1999; 990S-0137222.
PR 03-JUN-1999; 990S-0137528.
PR 04-JUN-1999; 990S-0137502.
PR 07-JUN-1999; 990S-0137724.
PR 08-JUN-1999; 990S-0138094.
PR 10-JUN-1999; 990S-0138540.
PR 14-JUN-1999; 990S-0138847.
PR 16-JUN-1999; 990S-0139119.
PR 17-JUN-1999; 990S-0139452.
PR 18-JUN-1999; 990S-0139453.
PR 18-JUN-1999; 990S-0139454.
PR 18-JUN-1999; 990S-0139455.
PR 18-JUN-1999; 990S-0139456.
PR 18-JUN-1999; 990S-0139457.
PR 18-JUN-1999; 990S-0139458.
PR 18-JUN-1999; 990S-0139459.
PR 18-JUN-1999; 990S-0139460.
PR 18-JUN-1999; 990S-0139461.
PR 18-JUN-1999; 990S-0139462.
PR 18-JUN-1999; 990S-0139463.
PR 18-JUN-1999; 990S-0139750.
PR 21-JUN-1999; 990S-0139817.
PR 22-JUN-1999; 990S-0139899.
PR 23-JUN-1999; 990S-0140353.
PR 23-JUN-1999; 990S-0140354.

PR 24-JUN-1999; 990S-0140695.
PR 28-JUN-1999; 990S-0140823.
PR 29-JUN-1999; 990S-0140991.
PR 30-JUN-1999; 990S-0141267.
PR 01-JUL-1999; 990S-0141842.
PR 01-JUL-1999; 990S-0142154.
PR 02-JUL-1999; 990S-0142055.
PR 06-JUL-1999; 990S-0142390.
PR 09-JUL-1999; 990S-0142803.
PR 12-JUL-1999; 990S-0142920.
PR 13-JUL-1999; 990S-0142977.
PR 14-JUL-1999; 990S-0143542.
PR 15-JUL-1999; 990S-0143624.
PR 16-JUL-1999; 990S-0144005.
PR 16-JUL-1999; 990S-0144085.
PR 19-JUL-1999; 990S-0144086.
PR 19-JUL-1999; 990S-0144325.
PR 19-JUL-1999; 990S-0144331.
PR 19-JUL-1999; 990S-0144332.
PR 19-JUL-1999; 990S-0144333.
PR 19-JUL-1999; 990S-0144334.
PR 20-JUL-1999; 990S-0144335.
PR 20-JUL-1999; 990S-0144352.
PR 20-JUL-1999; 990S-0144632.
PR 21-JUL-1999; 990S-0144684.
PR 21-JUL-1999; 990S-0144814.
PR 21-JUL-1999; 990S-0145086.
PR 22-JUL-1999; 990S-0145088.
PR 22-JUL-1999; 990S-0145089.
PR 22-JUL-1999; 990S-0145087.
PR 22-JUL-1999; 990S-0145192.
PR 23-JUL-1999; 990S-0145145.
PR 23-JUL-1999; 990S-0145218.
PR 26-JUL-1999; 990S-0145224.
PR 27-JUL-1999; 990S-0145276.
PR 27-JUL-1999; 990S-0145913.
PR 27-JUL-1999; 990S-0145918.
PR 28-JUL-1999; 990S-0145919.
PR 02-AUG-1999; 990S-0145951.
PR 02-AUG-1999; 990S-0146386.
PR 02-AUG-1999; 990S-0146388.
PR 02-AUG-1999; 990S-0146389.
PR 03-AUG-1999; 990S-0147038.
PR 04-AUG-1999; 990S-0147204.
PR 05-AUG-1999; 990S-0147302.
PR 05-AUG-1999; 990S-0147192.
PR 06-AUG-1999; 990S-0147260.
PR 06-AUG-1999; 990S-0147303.
PR 09-AUG-1999; 990S-0147416.
PR 09-AUG-1999; 990S-0147493.
PR 10-AUG-1999; 990S-0147935.
PR 11-AUG-1999; 990S-0148171.
PR 12-AUG-1999; 990S-0148319.
PR 13-AUG-1999; 990S-0148341.
PR 13-AUG-1999; 990S-0148565.
PR 16-AUG-1999; 990S-0148684.
PR 17-AUG-1999; 990S-0149368.
PR 18-AUG-1999; 990S-0149175.
PR 20-AUG-1999; 990S-0149426.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0149929.
PR 23-AUG-1999; 990S-0149902.
PR 23-AUG-1999; 990S-0149930.
PR 25-AUG-1999; 990S-0150566.
PR 26-AUG-1999; 990S-0150884.
PR 27-AUG-1999; 990S-0151065.
PR 27-AUG-1999; 990S-0151066.
PR 27-AUG-1999; 990S-0151080.
PR 30-AUG-1999; 990S-0151303.
PR 31-AUG-1999; 990S-0151438.
PR 01-SEP-1999; 990S-0151930.

```

PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0155596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match      62.5%; Score 15; DB 21; Length 1313;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 5 attatttggcctt 19
Db 1290 attatttggcctt 1304

```

```

RESULT 7
AA91624
ID AA91624 standard; DNA; 1392 BP.
AC AA91624;
DT 25-AUG-1999 (first entry)
DE Porphyromonas gingivalis protein PG7 ORF encoding DNA.
KW Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
KW vaccine; antigenic; ds.
OS Porphyromonas gingivalis.
XX MO9929870-A1.
XX 17-JUN-1999.
PD

```

```

XX 10-DEC-1998; 98WO-AU01023.
PF 04-AUG-1998; 98AU-0005028.
XX 10-DEC-1997; 97AU-0000839.
PR 31-DEC-1997; 97AU-0001182.
PR 30-JAN-1998; 98AU-0001546.
PR 10-MAR-1998; 98AU-0002264.
PR 09-APR-1998; 98AU-0002911.
PR 23-APR-1998; 98AU-0003128.
PR 05-MAY-1998; 98AU-0003338.
PR 22-MAY-1998; 98AU-0003654.
PR 29-JUL-1998; 98AU-0004917.
XX (CSLC-) CSL LTD.
XX Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
PI Ross BC, Rothel LJ, Webb EA;
XX WPT. 1999-385613/32.
DR P-PSDB; AA934406.
XX Antigenic Porphyromonas gingivalis peptides for preventing
PT gingivitis
PS Claim 12; Page 156; 588pp; English.
XX AA91536 to AA91801 encode two hundred and sixty six antigenic
CC Porphyromonas gingivalis (PG) polypeptide sequences given in AA934318 to
CC AA934583. AA91802 to AA91989 represent PCR primers used in the
CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
CC activity with a vaccine mechanism of action. The PG polypeptides can be
CC used as vaccines especially against Porphyromonas gingivalis. Probes can
CC be used to detect Porphyromonas gingivalis in standard hybridisation
CC assays. Porphyromonas gingivalis is involved in periodontal disease
XX especially gingivitis.
SQ Sequence 1392 BP; 361 A; 311 C; 328 G; 392 T; 0 other;

```

```

Query Match      62.5%; Score 15; DB 20; Length 1392;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 tcagattatttgg 15
Db 56 tcagattatttgg 70

```

```

RESULT 8
AAV33461
ID AAV33461 standard; cDNA; 1836 BP.
AC AAV33461;
DT 10-DEC-1998 (first entry)
DE Human T1-receptor ligand III splice variant 1 encoding cDNA.
XX Human T1-receptor ligand III; T1-R ligand III; atherosclerosis;
KW autoimmune disease; inflammation; metabolic dysfunction;
KW immune-regulated disorder; ss.
XX Homo sapiens.
OS
XX Key Location/Qualifiers
FH CDS 31..522
FT /*tag- a
FT /*product- "T1-R ligand III splice variant 1"
FT sig_peptide 31..102
FT /*tag- b
FT mat_peptide 103..519
FT /*tag- c

```

```

XX  MO9838311-A1.
XX
XX  03-SEP-1998.
XX
XX  26-FEB-1998; 98WO-US03483.
XX
XX  28-FEB-1997; 97US-0039483.
XX
XX  (HOMA-) HUMAN GENOME SCI INC.
XX
XX  Gentz, RL, Ni J, Ruben SM;
XX
XX  WPI: 1998-495398/42.
XX  P-PSDB; AAF70458.
XX
XX  New human T1-receptor ligand III splice variant(s) - useful for,
XX  e.g. detecting and treating immune system related disorders such as
XX  cancer and inflammation
XX
XX  Claim 2; Fig 1; 115pp; English.
XX
XX  The invention provides novel human T1-receptor ligand III (T1-R
XX  ligand III) splice variants and antibodies raised against these
XX  proteins. The present CDNA encodes the T1-R ligand III splice
XX  variant 1 protein. The T1-R ligand III splice variants are claimed
XX  useful for screening agonists and antagonists. They are also
XX  claimed useful for treating disorders such as atherosclerosis,
XX  autoimmune disease, inflammation, metabolic dysfunction and
XX  immune-regulated disorders.
XX
XX  Sequence 1836 BP; 637 A; 306 C; 353 G; 540 T; 0 other;

Query Match      62.5%; Score 15; DB 19; Length 1836;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  1 tcagattatttttg 15
    |||
Db  981 tcagattatttttg 995

RESULT  9
AAAF3852
ID  AAF93852 standard; CDNA; 2083 BP.
XX
XX  AAF93852;
XX
XX  23-MAY-2001 (first entry)
XX
XX  Human CDNA encoding a membrane or secretory protein clone PSEC0198.
XX
XX  Human: secretory protein; membrane protein; vaccine; gene therapy;
XX  Rheumatoid arthritis; diabetes; ss.
XX
XX  Homo sapiens.
XX
XX  EP1067182-A2.
XX
XX  10-JAN-2001.
XX
XX  07-JUL-2000; 2000EP-0114090.
XX
XX  08-JUL-1999; 99JP-0194179.
XX  11-JAN-2000; 2000JP-0118775.
XX  02-MAY-2000; 2000JP-0183766.
XX
XX  (HELI-) HELIX RES INST.
XX
XX  Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
XX  WPI: 2001-093989/11.

```

```

DR  P-PSDB; AAB88425.
XX
XX  Nucleic acids encoding secretory proteins/membrane proteins, useful in
XX  gene therapy or as candidate target molecules in drug development -
XX
XX  Claim 1; SEQ ID 217; 609pp + CD ROM; English.
XX
XX  This invention relates to nucleic acid sequences AAF93744 - AAF93916
XX  which encode human secretory or membrane proteins represented by
XX  AAB88317 - AAB88419. Included in the invention are primers
XX  AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
XX  CDNA sequences of the invention. The invention also includes methods for
XX  the production of antibodies directed against the proteins, and CDNA
XX  sequences, which can be used in vaccines. The polynucleotide sequences
XX  can be used in gene therapy. The polynucleotide sequences and the
XX  proteins they encode may be used in the prevention, treatment and
XX  diagnosis of diseases associated with inappropriate secretory
XX  protein/membrane protein expression. The nucleic acids and complementary
XX  sequences may also be used as DNA probes in diagnostic assays.
XX  (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
XX  presence of similar nucleic acid sequences in samples. They may also be
XX  used to study the expression and function of secretory proteins/membrane
XX  polypeptides and their role in metabolism. The polypeptides may be used
XX  as antigens in the production of antibodies against them and in assays to
XX  identify modulators (agonists and antagonists) of expression and
XX  activity. The antibodies and antagonists may also be used as therapeutic
XX  agents to down regulate expression and activity. The antibodies may also
XX  be used as diagnostic agents for detecting the presence of the
XX  polypeptides in samples (e.g. by enzyme linked immunosorbent assay
XX  (ELISA). Examples of diseases which may be treated include rheumatoid
XX  arthritis and diabetes.

XX  Sequence 2083 BP; 544 A; 467 C; 491 G; 581 T; 0 other;

Query Match      62.5%; Score 15; DB 22; Length 2083;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  1 tcagattatttttg 15
    |||
Db  1309 tcagattatttttg 1323

RESULT  10
AAA39981
ID  AAA39981 standard; CDNA; 2385 BP.
XX
XX  AAA39981;
XX
XX  16-OCT-2000 (first entry)
XX
XX  Human TANGO 187 CDNA.
XX
XX  TANGO 180; TANGO 181; TANGO 182; TANGO 183; TANGO 184; TANGO 185;
XX  TANGO 186; TANGO 188; TANGO 189; TANGO 215; TANGO 187; human, murine;
XX  secreted protein; transmembrane protein; gene therapy; vaccine;
XX  diagnosis; treatment; detection; ss.
XX
XX  Homo sapiens.
XX
XX  Key      Location/Qualifiers
XX  CDS      386..1135
XX           /*tag= a
XX           /product= "TANGO 187"

XX  WO200018904-A2.
XX
XX  06-APR-2000.
XX
XX  30-SEP-1999; 99WO-US22817.
XX
XX  30-SEP-1998; 98US-0164220.

```

PR 02-OCT-1998; 98US-0164169.
 XX
 PA (MILL-) MILLENNIUM BIOTHEAPUTICS INC.
 XX
 PI Barnes TM;
 XX
 DR WPI: 2000-293144/25.
 P-PSDB: AAB09923.
 XX
 XX Isolated nucleic acids encoding TANGO polypeptides useful for
 PT preventing, diagnosing and treating diseases associated with
 PT inappropriate protein expression
 XX
 PS Disclosure: Fig 52; 249pp; English.
 XX
 CC This invention describes novel human and murine nucleic acids encoding
 CC TANGO polypeptides (which are either wholly secreted or transmembrane
 CC proteins) which can be used for gene therapy and/or vaccination. The
 CC peptides are designated TANGO 180 to TANGO 189 and TANGO 215. The nucleic
 CC acids may be used to produce TANGO 180 to TANGO 189 and TANGO 215
 CC polypeptides according to standard recombinant DNA methodologies. They
 CC may also be used to detect and quantify the presence of TANGO nucleic
 CC acids in a sample and therefore identify or diagnose diseases associated
 CC with inappropriate TANGO expression (e.g. diseases related to over or
 CC under expression of the polypeptides or the expression of inactive
 CC polypeptides). The nucleic acids and the polypeptides they encode may be
 CC used according to standard gene therapy protocols, to treat diseases
 CC associated with inappropriate TANGO expression by supplementing a
 CC patients own production of the polypeptide of to rectify mutations that
 CC may result in expression of an abnormally active polypeptide. The
 CC antagonists of TANGO expression and activity which may be used to
 CC modulate TANGO related processes and diseases. The polypeptides are
 CC particularly useful for use as antigens for producing antibodies
 CC to TANGO proteins which may be used for inhibiting the activity of TANGO
 CC proteins. They may also be used to detect and quantify the presence of
 CC TANGO proteins in samples and therefore identify patients in whom the
 CC protein is over- or under-expressed. This sequence encodes the human
 CC TANGO 187 protein described in the method of the invention.
 XX
 S0 Sequence 2385 BP; 623 A; 519 C; 588 G; 653 T; 2 other;

Query Match 62.5%; Score 15; DB 21; Length 2385;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 tcagattatttgg 15
 |||
 Db 1279 tcagattatttgg 1293

RESULT 11
 AAA39953
 ID AAA39953 standard; CDNA; 2403 BP;
 XX
 AC AAA39953;
 XX
 DT 16-OCT-2000 (first entry)
 XX
 DE Human TANGO 187-1/3 cDNA.
 XX
 KW TANGO 180; TANGO 181; TANGO 182; TANGO 183; TANGO 184; TANGO 185;
 KW TANGO 186; TANGO 188; TANGO 189; TANGO 215; TANGO 187; human; murine;
 KW secreted protein; transmembrane protein; gene therapy; vaccine;
 KW diagnosis; treatment; detection; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 110..1141
 FT /*tag= a
 FT /product= "TANGO 187-1/3"

XX
 PN WO200018904-A2.
 XX
 PD 06-APR-2000.
 XX
 PF 30-SEP-1999; 99WO-US22817.
 XX
 PR 30-SEP-1998; 98US-0164220.
 XX
 PR 02-OCT-1998; 98US-0164169.
 XX
 PA (MILL-) MILLENNIUM BIOTHEAPUTICS INC.
 XX
 PI Barnes TM;
 XX
 DR WPI: 2000-293144/25.
 P-PSDB: AAB88281.
 XX
 XX Isolated nucleic acids encoding TANGO polypeptides useful for
 PT preventing, diagnosing and treating diseases associated with
 PT inappropriate protein expression
 XX
 PS Claim 1c; Fig 20; 249pp; English.
 XX
 CC This invention describes novel human and murine nucleic acids encoding
 CC TANGO polypeptides (which are either wholly secreted or transmembrane
 CC proteins) which can be used for gene therapy and/or vaccination. The
 CC peptides are designated TANGO 180 to TANGO 189 and TANGO 215. The nucleic
 CC acids may be used to produce TANGO 180 to TANGO 189 and TANGO 215
 CC polypeptides according to standard recombinant DNA methodologies. They
 CC may also be used to detect and quantify the presence of TANGO nucleic
 CC acids in a sample and therefore identify or diagnose diseases associated
 CC with inappropriate TANGO expression (e.g. diseases related to over or
 CC under expression of the polypeptides or the expression of inactive
 CC polypeptides). The nucleic acids and the polypeptides they encode may be
 CC used according to standard gene therapy protocols, to treat diseases
 CC associated with inappropriate TANGO expression by supplementing a
 CC patients own production of the polypeptide of to rectify mutations that
 CC may result in expression of an abnormally active polypeptide. The
 CC antagonists of TANGO expression and activity which may be used to
 CC modulate TANGO related processes and diseases. The polypeptides are
 CC particularly useful for use as antigens for producing antibodies
 CC to TANGO proteins which may be used for inhibiting the activity of TANGO
 CC proteins. They may also be used to detect and quantify the presence of
 CC TANGO proteins in samples and therefore identify patients in whom the
 CC protein is over- or under-expressed. This sequence encodes the human
 CC TANGO 187-1/3 protein described in the method of the invention.
 XX
 S0 Sequence 2403 BP; 660 A; 504 C; 566 G; 673 T; 0 other;

Query Match 62.5%; Score 15; DB 21; Length 2403;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 tcagattatttgg 15
 |||
 Db 1285 tcagattatttgg 1299

RESULT 12
 AAA39979
 ID AAA39979 standard; CDNA; 2418 BP.
 XX
 AC AAA39979;
 XX
 DT 16-OCT-2000 (first entry)
 XX
 DE Human TANGO 187-2 cDNA.
 XX
 KW TANGO 180; TANGO 181; TANGO 182; TANGO 183; TANGO 184; TANGO 185;
 KW TANGO 186; TANGO 188; TANGO 189; TANGO 215; TANGO 187; human; murine;
 KW secreted protein; transmembrane protein; gene therapy; vaccine;

KW diagnosis; treatment; detection; ss.
 XX Homo sapiens.
 XX
 XX
 FH Key Location/Qualifiers
 FT CDS 386..1168
 FT /*tag= a
 FT /product= "TANGO 187-2"
 XX
 PN MO200018904-A2.
 XX
 XX
 PD 06-APR-2000.
 XX
 XX
 PF 30-SEP-1999; 99WO-US22817.
 XX
 XX
 PR 30-SEP-1998; 98US-0164220.
 PR 02-OCT-1998; 98US-0164169.
 XX
 XX
 PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
 PI Barnes TM;
 XX
 DR WPI: 2000-293144/25.
 DR P-PSDB; AAY88295.
 XX
 XX
 PT Isolated nucleic acids encoding TANGO polypeptides useful for
 PT preventing, diagnosing and treating diseases associated with
 PT inappropriate protein expression
 XX
 XX
 PS Disclosure: Fig 50; 249pp; English.
 XX
 XX
 CC This invention describes novel human and murine nucleic acids encoding
 CC TANGO polypeptides (which are either wholly secreted or transmembrane
 CC proteins) which can be used for gene therapy and/or vaccination. The
 CC peptides are designated TANGO 180 to TANGO 189 and TANGO 215. The nucleic
 CC acids may be used to produce TANGO 180 to TANGO 189 and TANGO 215
 CC polypeptides according to standard recombinant DNA methodologies. They
 CC may also be used to detect and quantify the presence of TANGO nucleic
 CC acids in a sample and therefore identify or diagnose diseases associated
 CC with inappropriate TANGO expression (e.g. diseases related to over or
 CC under expression of the polypeptides and the expression of inactive
 CC polypeptides). The nucleic acids and the polypeptides they encode may be
 CC used according to standard gene therapy protocols, to treat diseases
 CC associated with inappropriate TANGO expression by supplementing a
 CC patients own production of the polypeptide of to rectify mutations that
 CC may result in expression of an abnormally active polypeptide. The
 CC polypeptides may also be used to identify and produce agonists and
 CC antagonists of TANGO expression and activity which may be used to
 CC modulate TANGO related processes and diseases. The polypeptides are
 CC particularly useful for use as antigens for producing antibodies
 CC to TANGO proteins which may be used for inhibiting the activity of TANGO
 CC proteins. They may also be used to detect and quantify the presence of
 CC TANGO proteins in samples and therefore identify patients in whom the
 CC protein is over- or under-expressed. This sequence encodes the human
 CC TANGO 187-2 protein described in the method of the invention.
 XX
 XX
 SO Sequence 2418 BP; 631 A; 528 C; 594 G; 663 T; 2 other;

Query Match 62.5%; Score 15; DB 21; Length 2418;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tcagattatttgg 15
 |||||
 DB 1312 tcagattatttgg 1326

RESULT 13
 AAA39975
 ID AAA39975 standard; cDNA; 2490 BP.
 XX
 AC AAA39975;

XX
 DT 16-OCT-2000 (first entry)
 XX
 XX
 DE Human TANGO 187-1 cDNA.
 XX
 XX
 KM TANGO 180; TANGO 181; TANGO 182; TANGO 183; TANGO 184; TANGO 185;
 KM TANGO 186; TANGO 188; TANGO 189; TANGO 215; TANGO 187; human; murine;
 KM secreted protein; transmembrane protein; gene therapy; vaccine;
 KM diagnosis; treatment; detection; ss.
 XX
 OS Homo sapiens.
 XX
 XX
 FH Key Location/Qualifiers
 FT CDS 386..1240
 FT /*tag= a
 FT /product= "TANGO 187-1"
 XX
 PN MO200018904-A2.
 XX
 XX
 PD 06-APR-2000.
 XX
 XX
 PF 30-SEP-1999; 99WO-US22817.
 XX
 XX
 PR 30-SEP-1998; 98US-0164220.
 PR 02-OCT-1998; 98US-0164169.
 XX
 XX
 PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
 PI Barnes TM;
 XX
 DR WPI: 2000-293144/25.
 DR P-PSDB; AAY88295.
 XX
 XX
 PT Isolated nucleic acids encoding TANGO polypeptides useful for
 PT preventing, diagnosing and treating diseases associated with
 PT inappropriate protein expression
 XX
 XX
 PS Disclosure: Fig 46; 249pp; English.
 XX
 XX
 CC This invention describes novel human and murine nucleic acids encoding
 CC TANGO polypeptides (which are either wholly secreted or transmembrane
 CC proteins) which can be used for gene therapy and/or vaccination. The
 CC peptides are designated TANGO 180 to TANGO 189 and TANGO 215. The nucleic
 CC acids may be used to produce TANGO 180 to TANGO 189 and TANGO 215
 CC polypeptides according to standard recombinant DNA methodologies. They
 CC may also be used to detect and quantify the presence of TANGO nucleic
 CC acids in a sample and therefore identify or diagnose diseases associated
 CC with inappropriate TANGO expression (e.g. diseases related to over or
 CC under expression of the polypeptides and the expression of inactive
 CC polypeptides). The nucleic acids and the polypeptides they encode may be
 CC used according to standard gene therapy protocols, to treat diseases
 CC associated with inappropriate TANGO expression by supplementing a
 CC patients own production of the polypeptide of to rectify mutations that
 CC may result in expression of an abnormally active polypeptide. The
 CC polypeptides may also be used to identify and produce agonists and
 CC antagonists of TANGO expression and activity which may be used to
 CC modulate TANGO related processes and diseases. The polypeptides are
 CC particularly useful for use as antigens for producing antibodies
 CC to TANGO proteins which may be used for inhibiting the activity of TANGO
 CC proteins. They may also be used to detect and quantify the presence of
 CC TANGO proteins in samples and therefore identify patients in whom the
 CC protein is over- or under-expressed. This sequence encodes the human
 CC TANGO 187-1 protein described in the method of the invention.
 XX
 XX
 SO Sequence 2490 BP; 642 A; 548 C; 626 G; 669 T; 5 other;

Query Match 62.5%; Score 15; DB 21; Length 2490;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tcagattatttgg 15
 |||||

DB 1384 tcagattatttgg 1398

RESULT 14
AAA39978
ID AAA39978 standard; cDNA: 2523 BP.
XX
AC AAA39978;
XX
DT 16-OCT-2000 (first entry)
XX
DE Human TANGO 187-1/2 CDNA.
XX
KW TANGO 180; TANGO 181; TANGO 182; TANGO 183; TANGO 184; TANGO 185;
KW TANGO 186; TANGO 188; TANGO 189; TANGO 215; TANGO 187; human; murine;
KW secreted protein; transmembrane protein; gene therapy; vaccine;
XX diagnosis; treatment; detection; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 386..1273
FT /tag= a "TANGO 187-1/2"
FT /product= "TANGO 187-1/2"
XX
PN WO200018904-A2.
XX
PD 06-APR-2000.
XX
PF 30-SEP-1999; 99WO-US22817.
XX
PR 30-SEP-1998; 98US-0164220.
PR 02-OCT-1998; 98US-0164169.
XX
XX (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
PI Barnes TM;
XX
DR WPI: 2000-293144/25.
DR P-PSDB: AAV88298.
XX
PT Isolated nucleic acids encoding TANGO polypeptides useful for
PT preventing, diagnosing and treating diseases associated with
XX inappropriate protein expression -
XX
PS Disclosure: Fig 49; 249pp; English.
XX
XX This invention describes novel human and murine nucleic acids encoding
CC TANGO polypeptides (which are either wholly secreted or transmembrane
CC proteins) which can be used for gene therapy and/or vaccination. The
CC peptides are designated TANGO 180 to TANGO 189 and TANGO 215. The nucleic
CC acids may be used to produce TANGO 180 to TANGO 189 and TANGO 215
CC polypeptides according to standard recombinant DNA methodologies. They
CC may also be used to detect and quantify the presence of TANGO nucleic
CC acids in a sample and therefore identify or diagnose diseases associated
CC with inappropriate TANGO expression (e.g. diseases related to over or
CC under expression) of the polypeptides or the expression of inactive
CC polypeptides). The nucleic acids and the polypeptides they encode may be
CC used according to standard gene therapy protocols, to treat diseases
CC associated with inappropriate TANGO expression by supplementing a
CC patient's own production of the polypeptide of to rectify mutations that
CC may result in expression of an abnormally active polypeptide. The
CC polypeptides may also be used to identify and produce agonists and
CC antagonists of TANGO expression and activity which may be used to
CC modulate TANGO related processes and diseases. The polypeptides are
CC particularly useful for use as antigens for producing antibodies
CC to TANGO proteins which may be used for inhibiting the activity of TANGO
CC proteins. They may also be used to detect and quantify the presence of
CC TANGO proteins in samples and therefore identify patients in whom the
CC protein is over- or under-expressed. This sequence encodes the human
CC TANGO 187-1/2 protein described in the method of the invention.
SQ Sequence 2523 BP; 650 A; 557 C; 632 G; 679 T; 5 other;

Query Match 62.5%; Score 15; DB 21; Length 2523;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcagattatttgg 15
|||||
DB 1417 tcagattatttgg 1431

RESULT 15
AAA39980
ID AAA39980 standard; cDNA: 2562 BP.
XX
AC AAA39980;
XX
DT 16-OCT-2000 (first entry)
XX
DE Human TANGO 187-3 CDNA.
XX
KW TANGO 180; TANGO 181; TANGO 182; TANGO 183; TANGO 184; TANGO 185;
KW TANGO 186; TANGO 188; TANGO 189; TANGO 215; TANGO 187; human; murine;
KW secreted protein; transmembrane protein; gene therapy; vaccine;
XX diagnosis; treatment; detection; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 386..1312
FT /tag= a "TANGO 187-3"
FT /product= "TANGO 187-3"
XX
PN WO200018904-A2.
XX
PD 06-APR-2000.
XX
PF 30-SEP-1999; 99WO-US22817.
XX
PR 30-SEP-1998; 98US-0164220.
PR 02-OCT-1998; 98US-0164169.
XX
XX (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
PI Barnes TM;
XX
DR WPI: 2000-293144/25.
DR P-PSDB: AAV88300.
XX
PT Isolated nucleic acids encoding TANGO polypeptides useful for
PT preventing, diagnosing and treating diseases associated with
XX inappropriate protein expression -
XX
PS Disclosure: Fig 51; 249pp; English.
XX
XX This invention describes novel human and murine nucleic acids encoding
CC TANGO polypeptides (which are either wholly secreted or transmembrane
CC proteins) which can be used for gene therapy and/or vaccination. The
CC peptides are designated TANGO 180 to TANGO 189 and TANGO 215. The nucleic
CC acids may be used to produce TANGO 180 to TANGO 189 and TANGO 215
CC polypeptides according to standard recombinant DNA methodologies. They
CC may also be used to detect and quantify the presence of TANGO nucleic
CC acids in a sample and therefore identify or diagnose diseases associated
CC with inappropriate TANGO expression (e.g. diseases related to over or
CC under expression) of the polypeptides or the expression of inactive
CC polypeptides). The nucleic acids and the polypeptides they encode may be
CC used according to standard gene therapy protocols, to treat diseases
CC associated with inappropriate TANGO expression by supplementing a
CC patient's own production of the polypeptide of to rectify mutations that
CC may result in expression of an abnormally active polypeptide. The
CC polypeptides may also be used to identify and produce agonists and
CC antagonists of TANGO expression and activity which may be used to
CC modulate TANGO related processes and diseases. The polypeptides are

CC particularly useful for use as antigens for producing antibodies
CC to TANGO proteins which may be used for inhibiting the activity of TANGO
CC proteins. They may also be used to detect and quantify the presence of
CC TANGO proteins in samples and therefore identify patients in whom the
CC protein is over- or under-expressed. This sequence encodes the human
CC TANGO 187-3 protein described in the method of the invention.
XX

SQ Sequence 2562 BP; 672 A; 560 C; 626 G; 701 T; 3 other;

Query Match 62.5%; Score 15; DB 21; Length 2562;
Best Local Similarity 100.0%; Pred. No. 15;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tcagattatttgg 15
|||||
DB 1456 tcagattatttgg 1470

Search completed: November 15, 2001, 07:55:57
Job time: 7257 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 15, 2001, 04:03:29 ; Search time 49.19 Seconds

(Without alignments)
110.500 Million cell updates/sec

Title: US-09-663-020-7
Perfect score: 24
Sequence: 1 tcagattatttggttcacac 24

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 residues

Word size : 0

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/ptodata/2/1na/5A.COMB.seq:*
- 2: /cgn2_6/ptodata/2/1na/5B.COMB.seq:*
- 3: /cgn2_6/ptodata/2/1na/6A.COMB.seq:*
- 4: /cgn2_6/ptodata/2/1na/6B.COMB.seq:*
- 5: /cgn2_6/ptodata/2/1na/PCITUS.COMB.seq:*
- 6: /cgn2_6/ptodata/2/1na/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	15	62.5	1303 2	US-08-793-410-11
2	14	58.3	486 2	US-08-506-864A-5
3	14	58.3	486 2	US-08-851-968-5
4	14	58.3	1100 3	US-09-484-345-10
5	14	58.3	1502 2	US-08-506-864A-2
6	14	58.3	1502 2	US-08-851-968-2
7	14	58.3	2539 3	US-08-788-231A-1
8	14	58.3	3282 4	US-09-315-793-51
9	14	58.3	4673 5	US-07-638-431-1
10	14	58.3	4673 5	PCT-US92-00018-1
11	13	54.2	23 2	US-08-479-275D-33
12	13	54.2	23 2	US-08-488-271B-33
13	13	54.2	42 3	US-08-631-319A-3
14	13	54.2	42 4	US-08-842-306B-17
15	13	54.2	42 4	US-08-838-973B-15
16	13	54.2	467 4	US-09-253-584-8
17	13	54.2	876 5	PCT-US96-05320A-1122
18	13	54.2	1520 3	US-09-100-664A-1
19	13	54.2	2215 2	US-08-980-329C-1
20	13	54.2	2442 4	US-09-040-485-1
21	13	54.2	2539 3	US-08-581-148C-20
22	13	54.2	2797 2	US-08-555-723B-4
23	13	54.2	2797 3	US-09-123-465-4
24	13	54.2	2875 3	US-08-714-918-8
25	13	54.2	2875 4	US-09-265-315-8
26	13	54.2	2875 4	US-09-265-315-8
27	13	54.2	2875 4	US-09-266-417-8

28	13	54.2	3012 2	US-08-475-427-5	Sequence 5, Appl
29	13	54.2	3012 2	US-07-842-165-5	Sequence 5, Appl
30	13	54.2	3268 3	US-09-356-952-13	Sequence 13, Appl
31	13	54.2	3515 2	US-08-391-743A-1	Sequence 1, Appl
32	13	54.2	3892 2	US-08-555-723B-3	Sequence 3, Appl
33	13	54.2	3892 2	US-09-123-465-3	Sequence 3, Appl
34	13	54.2	4291 2	US-08-417-210A-80	Sequence 80, Appl
35	13	54.2	4857 2	US-08-566-398-16	Sequence 16, Appl
36	13	54.2	5561 2	US-08-400-159-1	Sequence 1, Appl
37	13	54.2	5561 3	US-08-611-729A-1	Sequence 1, Appl
38	13	54.2	6519 1	US-08-588-985-1	Sequence 1, Appl
39	13	54.2	6519 1	US-08-971-988-1	Sequence 1, Appl
40	13	54.2	6628 3	US-08-815-809-3	Sequence 3, Appl
41	13	54.2	6649 2	US-08-816-155B-5	Sequence 5, Appl
42	13	54.2	6649 3	US-09-079-587-5	Sequence 5, Appl
43	13	54.2	7091 2	US-08-658-665-40	Sequence 40, Appl
44	13	54.2	7091 4	US-08-796-101-4	Sequence 4, Appl
45	13	54.2	7091 4	US-09-085-273-40	Sequence 40, Appl

ALIGNMENTS

RESULT 1
US-08-793-410-11
Sequence 11, Application US/08793410
Patent No. 5955650
GENERAL INFORMATION:
APPLICANT: HITZ, WILLIAM DEAN
TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF CANOLA
AND SOYBEAN PALMITOYL-ACP THIO-
ESTERASE GENES AND THEIR USE IN
THE REGULATION OF FATTY ACID
CONTENT OF THE OILS OF SOYBEAN
NUMBER OF INVENTION: 32
CORRESPONDENCE ADDRESSES:
ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: USA
ZIP: 19898
COMPUTER READABLE FORM: 3.50 INCH
MEDIUM TYPE: DISKETTE
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MICROSOFT WINDOWS 95
SOFTWARE: MICROSOFT WORD VERSION 7.0A
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/793,410
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10627
FILING DATE: AUGUST 25, 1995
ATTORNEY/AGENT INFORMATION:
NAME: CHRISTENBURY, LYNN M.
REGISTRATION NUMBER: 30,971
REFERENCE/DOCKET NUMBER: CR-9567-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-773-0164
TELEFAX: 302-773-0164
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1303 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-793-410-11

Query Match 62.5%; Score 15; DB 2; Length 1303;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ttatttggttc 20
|||||
DB 1103 TTTATTGCGCTTC 1117

RESULT 2

US-08-506-864A-5/C
; Sequence 5, Application US/08506864A
; Patent No. 5834245

GENERAL INFORMATION:

APPLICANT: NAKAMURA, YUSUKE
APPLICANT: FUJIMARA, YOSHIYUKI
TITLE OF INVENTION: PRITS PROTEINS AND DNA'S
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLYNN, THEIL, BOUTELL & TANIS, P.C.
STREET: 2026 Rambling Road
CITY: Kalamazoo
STATE: Michigan
COUNTRY: USA
ZIP: 49008-1699

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inches, 1.44 Mb storage
COMPUTER: IBM PC/XT/AT Compatible
OPERATING SYSTEM: MS-DOS 5.0
SOFTWARE: WordPerfect 5.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/506,864A
FILING DATE:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP6-178131
FILING DATE: 29-JULY-1994

ATTORNEY/AGENT INFORMATION:
NAME: Terryence F. Chapman
REGISTRATION NUMBER: 32549

REFERENCE/DOCKET NUMBER: Furuya Case 1334
TELECOMMUNICATION INFORMATION:
TELEPHONE: (616) 381-1156
TELEFAX: (616) 381-5465

INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 486

TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens

IMMEDIATE SOURCE:
LIBRARY: human DNA cosmid library

FEATURE:
NAME/KEY: exon 3
LOCATION: 74..371

IDENTIFICATION METHOD: experimental examination
US-08-506-864A-5

Query Match 58.3%; Score 14; DB 2; Length 486;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 atttggttcac 22
|||||
DB 164 ATTTGGGCTTCAC 151

RESULT 3
US-08-851-968-5/C
; Sequence 5, Application US/08851968
; Patent No. 5935786

GENERAL INFORMATION:

APPLICANT: NAKAMURA, YUSUKE
APPLICANT: FUJIMARA, YOSHIYUKI
TITLE OF INVENTION: PRITS PROTEINS AND DNA'S
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLYNN, THEIL, BOUTELL & TANIS, P.C.
STREET: 2026 Rambling Road
CITY: Kalamazoo
STATE: Michigan
COUNTRY: USA
ZIP: 49008-1699

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inches, 1.44 Mb storage
COMPUTER: IBM PC/XT/AT Compatible
OPERATING SYSTEM: MS-DOS 5.0
SOFTWARE: WordPerfect 5.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/851,968
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/506,864
FILING DATE:

APPLICATION NUMBER: JP6-178131
FILING DATE: 29-JULY-1994

ATTORNEY/AGENT INFORMATION:
NAME: Terryence F. Chapman
REGISTRATION NUMBER: 32549
REFERENCE/DOCKET NUMBER: Furuya Case 1334
TELECOMMUNICATION INFORMATION:
TELEPHONE: (616) 381-1156
TELEFAX: (616) 381-5465

INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 486

TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens

IMMEDIATE SOURCE:
LIBRARY: human DNA cosmid library

FEATURE:
NAME/KEY: exon 3
LOCATION: 74..371

IDENTIFICATION METHOD: experimental examination
US-08-851-968-5

Query Match 58.3%; Score 14; DB 2; Length 486;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 atttggttcac 22
|||||
DB 164 ATTTGGGCTTCAC 151

RESULT 4

US-09-484-345-10
; Sequence 10, Application US/09484345
; Patent No. 6159734

GENERAL INFORMATION:

APPLICANT: Robert McKay
APPLICANT: Alexander H. Borchers
APPLICANT: Brenda F. Baker

TITLE OF INVENTION: ANTISENSE MODULATION OF PEROXISOME PROLIFERATOR-ACTIVATED RECEPTOR
 FILE REFERENCE: RTS-0104
 CURRENT APPLICATION NUMBER: US/09/484,345
 CURRENT FILING DATE: 2000-01-18
 NUMBER OF SEQ ID NOS: 90
 SEQ ID NO 10
 LENGTH: 1100
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (859)...(940)
 US-09-484-345-10

Query Match 58.3%; Score 14; DB 3; Length 1100;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 atttggccttcac 22
 Db 218 atttggccttcac 231

RESULT 5
 US-08-506-864A-2/C
 Sequence 2, Application US/08506864A
 Patent No. 5834245
 GENERAL INFORMATION:
 APPLICANT: NAKAMURA, YUSUKE
 APPLICANT: FUJIMARA, YOSHIYUKI
 TITLE OF INVENTION: PRITS PROTEINS AND DNA'S
 TITLE OF INVENTION: ENCODING THE SAME
 NUMBER OF SEQUENCES: 21
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: FLYNN, THIEL, BOUTELL & TANIS, P.C.
 STREET: 2026 Rambling Road
 CITY: Kalamazoo
 STATE: Michigan
 COUNTRY: USA
 ZIP: 49008-1699
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.5 inches, 1.44 Mb storage
 COMPUTER: IBM PC/XT/AT Compatible
 OPERATING SYSTEM: MS-DOS 5.0
 SOFTWARE: WordPerfect 5.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/506,864A
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP6-178131
 FILING DATE: 29-JULY-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Terrence F. Chapman
 REGISTRATION NUMBER: 32549
 REFERENCE/DOCKET NUMBER: Furuya Case 1334
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (616) 381-1156
 TELEFAX: (616) 381-5465
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1502
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA to mRNA
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 IMMEDIATE SOURCE:
 LIBRARY: human fetal lung cDNA library
 FEATURE:
 NAME/KEY: CDS

LOCATION: 62,1189
 IDENTIFICATION METHOD: experimental examination
 US-08-506-864A-2

Query Match 58.3%; Score 14; DB 2; Length 1502;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 atttggccttcac 22
 Db 207 atttggccttcac 194

RESULT 6
 US-08-851-968-2/C
 Sequence 2, Application US/08851968
 Patent No. 5935786
 GENERAL INFORMATION:
 APPLICANT: NAKAMURA, YUSUKE
 APPLICANT: FUJIMARA, YOSHIYUKI
 TITLE OF INVENTION: PRITS PROTEINS AND DNA'S
 TITLE OF INVENTION: ENCODING THE SAME
 NUMBER OF SEQUENCES: 21
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: FLYNN, THIEL, BOUTELL & TANIS, P.C.
 STREET: 2026 Rambling Road
 CITY: Kalamazoo
 STATE: Michigan
 COUNTRY: USA
 ZIP: 49008-1699
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.5 inches, 1.44 Mb storage
 COMPUTER: IBM PC/XT/AT Compatible
 OPERATING SYSTEM: MS-DOS 5.0
 SOFTWARE: WordPerfect 5.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/851,968
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/506,864
 FILING DATE:
 APPLICATION NUMBER: JP6-178131
 FILING DATE: 29-JULY-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Terrence F. Chapman
 REGISTRATION NUMBER: 32549
 REFERENCE/DOCKET NUMBER: Furuya Case 1334
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (616) 381-1156
 TELEFAX: (616) 381-5465
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1502
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA to mRNA
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 IMMEDIATE SOURCE:
 LIBRARY: human fetal lung cDNA library
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 62,1189
 IDENTIFICATION METHOD: experimental examination
 US-08-851-968-2

Query Match 58.3%; Score 14; DB 2; Length 1502;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 9 atttggctcac 22
Db 207 ATTtGGGCTTCA 194

RESULT 7

US-08-788-231A-1/C
; Sequence 1, Application US/08788231A
; Patent No. 6019974
; GENERAL INFORMATION:
; APPLICANT: L'Hernault, Steven W.
; TITLE OF INVENTION: SPE-4 RELATED PEPTIDES, ANTIBODIES AND
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/788,231A
; FILING DATE: 24-JAN-1997
; CLASSIFICATION: 424
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 60/010,672
; FILING DATE: 26-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Feibert, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 60-95
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2539 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: not relevant
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: *Caenorhabditis elegans*
; IMMEDIATE SOURCE:
; CLONE: SPE-4
; US-08-788-231A-1

Query Match 58.3%; Score 14; DB 3; Length 2539;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 tcaattattttg 14
Db 1478 TCAGATTATTATTG 1465

RESULT 8

US-09-315-793-51/C
; Sequence 51, Application US/09315793
; Patent No. 6221597
; GENERAL INFORMATION:
; APPLICANT: Roberts, Christopher J.
; TITLE OF INVENTION: ESSENTIAL GENES OF YEAST AS TARGETS FOR ANTIFUNGAL
; TITLE OF INVENTION: AGENTS, HERBICIDES, INSECTICIDES AND ANTI-PROLIFERATION

; TITLE OF INVENTION: DRUGS
; FILE REFERENCE: 9301-048
; CURRENT APPLICATION NUMBER: US/09/315,793
; CURRENT FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 51
; LENGTH: 3282
; TYPE: DNA
; ORGANISM: *Saccharomyces cerevisiae*
; US-09-315-793-51

Query Match 58.3%; Score 14; DB 4; Length 3282;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 talttggctca 21
Db 2374 TATTGGGCTTCA 2361

RESULT 9

US-07-638-431-1/C
; Sequence 1, Application US/07638431
; Patent No. 5198535
; GENERAL INFORMATION:
; APPLICANT: Hoffman, Stephen L.
; APPLICANT: Charoenyit, Yupin
; APPLICANT: Hedstrom, Richard
; APPLICANT: Khusmith, Srisid
; APPLICANT: Rogers IV, William O.
; TITLE OF INVENTION: Protective malaria sporozoite surface protein
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: A. David Spevack
; STREET: NMRDC Building 1 T-12 National Naval
; STREET: Medical Center
; CITY: Bethesda
; STATE: MD
; COUNTRY: USA
; ZIP: 20814-5044
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/638,431
; FILING DATE: 19910110
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Spevack, Avrom D.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 295-6759
; TELEFAX: (301) 295-4033
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4673 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: N
; ANTI-SENSE: N
; ORIGINAL SOURCE:
; ORGANISM: *Plasmodium yoelii*
; STRAIN: 17X(NL)
; DEVELOPMENTAL STAGE: erythrocytic stage
; TISSUE TYPE: Blood
; CELL TYPE: erythrocytic stage
; IMMEDIATE SOURCE:

LIBRARY: Py-lambda9t11-2-7 kb genomic expression
CLONE: Py10.1111
FEATURE:
NAME/KEY: CDS
LOCATION: 718..3195
OTHER INFORMATION:
US-07-638-431-1

Query Match
Best Local Similarity 100.0%; Score 14; DB 1; Length 4673;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 tatttggcttca 21
|||||
DB 1375 TATTTGGCTTCA 1362

RESULT 10

PCT-US92-00018-1/c
; Sequence 1, Application PC/TUS9200018
; GENERAL INFORMATION:
; APPLICANT: Hoffman, Stephen L.
; APPLICANT: Charoenvit, Yupin
; APPLICANT: Hedstrom, Richard
; APPLICANT: Khumsmith, Srisin
; APPLICANT: Rogers IV, William O.
; TITLE OF INVENTION: Protective malaria sporozoite surface protein
; NUMBER OF INVENTION: Immunogen and gene encoding
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: A. David Spevack
; STREET: NMRDC Building 1 T-12 National Naval
; STREET: Medical Center
; CITY: Bethesda
; STATE: MD
; COUNTRY: USA
; ZIP: 20814-5044
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/00018
; FILING DATE: 19920103
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Spevack, Avram D.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 295-6759
; TELEFAX: (301) 295-4033
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4673 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: N
; ANTI-SENSE: N
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium yoelii
; STRAIN: 17X(NL)
; DEVELOPMENTAL STAGE: erythrocytic stage
; TISSUE TYPE: Blood
; CELL TYPE: erythrocytic stage
; IMMEDIATE SOURCE:
; LIBRARY: Py-lambda9t11-2-7 kb genomic expression
; CLONE: Py10.1111
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 718..3195

OTHER INFORMATION:
PCT-US92-00018-1

Query Match
Best Local Similarity 100.0%; Score 14; DB 5; Length 4673;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 tatttggcttca 21
|||||
DB 1375 TATTTGGCTTCA 1362

RESULT 11

US-08-479-275D-33/c
; Sequence 33, Application US/08479275D
; Patent No. 5869438
; GENERAL INFORMATION:
; APPLICANT: Svendsen, Allan
; APPLICANT: Packar, Shankant A.
; APPLICANT: Gormsen, Erik
; APPLICANT: Clausen, Ib G.
; APPLICANT: Okkels, Jens S.
; APPLICANT: Thelersen, Marianne
; TITLE OF INVENTION: LIPASE VARIANTS
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 58694380 No. 58694380disk of No. 5869438th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10174-6201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/479,275D
; FILING DATE: 07-JUNE-1995
; CLASSIFICATION: 510
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3520,514-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 23 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-479-275D-33

Query Match
Best Local Similarity 100.0%; Score 13; DB 2; Length 23;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 ttgggcttact 23
|||||
DB 14 TTTGGCTTACT 2

RESULT 12
US-08-488-271B-33/c
; Sequence 33, Application US/08488271B
; Patent No. 5892013
; GENERAL INFORMATION:

APPLICANT: Svendsen, Allan
APPLICANT: Palkar, Shamkant A.
APPLICANT: Gormsen, Erik
APPLICANT: Clausen, Ib G.
APPLICANT: Okels, Jens S.
APPLICANT: Thellersen, Marianne
TITLE OF INVENTION: LIPASE VARIANTS
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 58920130 No. 5892013disk of No. 5892013th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174-6201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,271B
FILING DATE: 07-JUNE-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3520.504-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-488-271B-33

Query Match 54.2%; Score 13; DB 2; Length 23;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 ttgggcttcaact 23
11 tttgggcttcaact 23
DB 14 tttgggcttcaact 2

RESULT 13
US-08-631-319A-3
Sequence 3, Application US/08631319A
Patent No. 6117641
GENERAL INFORMATION:
APPLICANT: Berlin, Vivian
APPLICANT: Levin, David
APPLICANT: Ohya, Yoshikazu
TITLE OF INVENTION: Assay and Reagents for Identifying Anti-
TITLE OF INVENTION: Fungal Agents, and Uses Related Thereto
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY, HOAG & ELIOT LLP
STREET: ONE POST OFFICE SQUARE
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII (text)

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/631,319A
FILING DATE: 11-APR-1998
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Pham, Chinh H.
REGISTRATION NUMBER: 39,329
REFERENCE/DOCKET NUMBER: MIV-074.02
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 832-1000
TELEFAX: (617) 832-7000
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 42 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Other nucleic acid
ANTI-SENSE: Yes
US-08-631-319A-3

Query Match 54.2%; Score 13; DB 3; Length 42;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 cagattatttgg 14
11 tttgggcttcaact 23
DB 30 cagattatttgg 42

RESULT 14
US-08-842-306B-17
Sequence 17, Application US/08842306B
Patent No. 6271197
GENERAL INFORMATION:
APPLICANT: Berlin, Vivian
APPLICANT: Levin, David
APPLICANT: Ohya, Yoshikazu
APPLICANT: Damagnez, Veronique
APPLICANT: Smith, Susan
TITLE OF INVENTION: ASSAYS AND REAGENTS FOR IDENTIFYING
TITLE OF INVENTION: ANTI-FUNGAL AGENTS, AND USES RELATED THERETO
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY, HOAG & ELIOT LLP
STREET: ONE POST OFFICE SQUARE
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109-2170
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPad
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/842,306B
FILING DATE: 23-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/771,212
FILING DATE: 20-DEC-1996
APPLICATION NUMBER: US 08/631,319
FILING DATE: 11-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: MIV-074.04
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:

LENGTH: 42 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: other nucleic acid
 DESCRIPTION: /desc = "primer"
 SEQUENCE DESCRIPTION: SEQ ID NO: 17:
 US-08-842-306B-17

DB 30 CAGATTATTG 42

Search completed: November 15, 2001, 05:55:52
 Job time: 6743 sec

Query Match 54.2%; Score 13; DB 4; Length 42;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 cagattattg 14
 DB 30 CAGATTATTG 42

RESULT 15
 US-08-838-973B-15
 ; Sequence 15, Application US/08838973B
 ; Patent No. 6277564
 ; GENERAL INFORMATION:
 ; APPLICANT: Berlin, Vivian
 ; Damaguez, Veronique
 ; Smith, Susan
 ; TITLE OF INVENTION: ASSAYS AND REAGENTS FOR IDENTIFYING
 ; ANTI-FUNGAL AGENTS, AND USES RELATED THEREO
 ; NUMBER OF SEQUENCES: 48
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: FOLEY, HOAG & ELLIOT LLP
 ; STREET: One Post Office Square
 ; CITY: Boston
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 02109-2170
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/838,973B
 ; FILING DATE: 23-Apr-1997
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/631,319
 ; FILING DATE: 10-APR-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Vincent, Matthew P.
 ; REGISTRATION NUMBER: 36,709
 ; REFERENCE/DOCKET NUMBER: MIV-074.05
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617-832-1000
 ; TELEFAX: 617-832-7000
 ; INFORMATION FOR SEQ ID NO: 15:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 42 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: other nucleic acid
 ; DESCRIPTION: /desc = "primer"
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 15:
 US-08-838-973B-15

Query Match 54.2%; Score 13; DB 4; Length 42;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 cagattattg 14
 DB 30 CAGATTATTG 42

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 15, 2001, 04:04:24 : Search time 1315.37 Seconds
(without alignments)
172.475 Million cell updates/sec

Title: US-09-663-020-7
Perfect score: 24
Sequence: 1 tcagattatttggtcactc 24

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 10228115 seqs, 4726426750 residues

Word size : 0

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*

1:	gb_est1:*
2:	gb_est2:*
3:	gb_est3:*
4:	gb_est4:*
5:	gb_est5:*
6:	gb_est6:*
7:	gb_est7:*
8:	gb_est8:*
9:	gb_est9:*
10:	gb_est10:*
11:	gb_est11:*
12:	gb_est12:*
13:	gb_est13:*
14:	gb_est14:*
15:	gb_est15:*
16:	gb_est16:*
17:	gb_est17:*
18:	gb_est18:*
19:	gb_est19:*
20:	gb_est20:*
21:	gb_est21:*
22:	gb_est22:*
23:	gb_est23:*
24:	gb_est24:*
25:	gb_est25:*
26:	gb_est26:*
27:	gb_est27:*
28:	gb_est28:*
29:	gb_est29:*
30:	gb_est30:*
31:	gb_est31:*
32:	gb_est32:*
33:	gb_est33:*
34:	gb_est34:*
35:	gb_est35:*
36:	gb_est36:*
37:	gb_est37:*
38:	gb_est38:*
39:	gb_est39:*
40:	gb_est40:*
41:	gb_est41:*
42:	gb_est42:*
43:	gb_est43:*
44:	gb_est44:*
45:	gb_est45:*
46:	gb_est46:*
47:	gb_est47:*
48:	gb_est48:*
49:	gb_est49:*
50:	gb_est50:*
51:	gb_est51:*
52:	gb_est52:*
53:	gb_est53:*
54:	gb_est54:*
55:	gb_est55:*
56:	gb_est56:*
57:	gb_est57:*
58:	gb_est58:*
59:	gb_est59:*
60:	gb_est60:*
61:	gb_est61:*
62:	gb_est62:*
63:	gb_est63:*
64:	gb_est64:*
65:	gb_est65:*
66:	gb_est66:*
67:	gb_est67:*
68:	gb_est68:*
69:	gb_est69:*
70:	gb_est70:*
71:	gb_est71:*
72:	gb_est72:*
73:	gb_est73:*
74:	gb_est74:*
75:	gb_est75:*
76:	gb_est76:*
77:	gb_est77:*
78:	gb_est78:*
79:	gb_est79:*
80:	gb_est80:*
81:	gb_est81:*
82:	gb_est82:*
83:	gb_est83:*
84:	gb_est84:*
85:	gb_est85:*
86:	gb_est86:*
87:	gb_est87:*
88:	gb_est88:*
89:	gb_est89:*
90:	gb_est90:*
91:	gb_est91:*
92:	gb_est92:*
93:	gb_est93:*
94:	gb_est94:*
95:	gb_est95:*
96:	gb_est96:*
97:	gb_est97:*
98:	gb_est98:*
99:	gb_est99:*
100:	gb_est100:*
101:	gb_est101:*
102:	gb_est102:*
103:	gb_est103:*
104:	gb_est104:*
105:	gb_est105:*
106:	gb_est106:*
107:	gb_est107:*
108:	gb_est108:*
109:	gb_est109:*
110:	gb_est110:*
111:	gb_est111:*
112:	gb_est112:*
113:	gb_est113:*
114:	gb_est114:*
115:	gb_est115:*
116:	gb_est116:*


```
117: gb_est148:*
118: gb_est149:*
119: gb_est150:*
120: gb_est151:*
121: gb_est152:*
122: gb_est153:*
123: gb_est154:*
124: gb_est155:*
125: gb_est156:*
126: gb_est157:*
127: gb_est158:*
128: gb_est159:*
129: gb_est160:*
130: gb_est161:*
131: gb_est162:*
132: gb_est163:*
133: gb_est164:*
134: gb_est165:*
135: gb_est166:*
136: gb_est167:*
137: gb_est168:*
138: gb_est169:*
139: gb_est170:*
140: gb_est171:*
141: gb_est172:*
142: gb_est173:*
143: gb_est174:*
144: gb_est175:*
145: gb_est176:*
146: gb_est177:*
147: gb_est178:*
148: gb_est179:*
149: gb_est180:*
150: gb_est181:*
151: gb_est182:*
152: gb_est183:*
153: gb_est184:*
154: gb_est185:*
155: gb_est186:*
156: gb_est187:*
157: gb_est188:*
158: gb_est189:*
159: gb_est190:*
160: gb_est191:*
161: gb_est192:*
162: gb_est193:*
163: gb_est194:*
164: gb_est195:*
165: gb_est196:*
166: gb_est197:*
167: gb_est198:*
168: gb_est199:*
169: gb_est200:*
170: gb_est201:*
171: gb_est202:*
172: gb_est203:*
173: gb_est204:*
174: gb_est205:*
175: gb_est206:*
176: gb_est207:*
177: gb_est208:*
178: gb_est209:*
179: gb_est210:*
180: gb_est211:*
181: gb_est212:*
182: gb_est213:*
183: gb_est214:*
184: gb_est215:*
185: gb_est216:*
186: gb_est217:*
187: gb_est218:*
188: gb_est219:*
189: gb_est220:*

190: gb_est110:*
191: gb_est111:*
192: gb_est112:*
193: gb_est113:*
194: gb_est114:*
195: gb_est115:*
196: gb_est116:*
197: gb_est117:*
198: gb_est118:*
199: gb_est119:*
200: gb_est120:*
201: gb_est121:*
202: gb_est122:*
203: gb_est123:*
204: gb_est124:*
205: gb_est125:*
206: gb_est126:*
207: gb_est127:*
208: gb_est128:*
209: gb_est129:*
210: gb_est130:*
211: gb_est131:*
212: gb_est132:*
213: gb_est133:*
214: gb_est134:*
215: gb_est135:*
216: gb_est136:*
217: gb_est137:*
218: gb_est138:*
219: gb_est139:*
220: gb_est140:*
221: gb_est141:*
222: gb_est142:*
223: gb_est143:*
224: gb_est144:*
225: gb_est145:*
226: gb_est146:*
227: gb_est147:*
228: gb_est148:*
229: gb_est149:*
230: gb_est150:*
231: gb_est151:*
232: gb_est152:*
233: gb_est153:*
234: gb_est154:*
235: gb_est155:*
236: gb_est156:*
237: gb_est157:*
238: gb_est158:*
239: gb_est159:*
240: gb_est160:*
241: gb_est161:*
242: gb_est162:*
243: gb_est163:*
244: gb_est164:*
245: gb_est165:*
246: gb_est166:*
247: gb_est167:*
248: gb_est168:*
249: gb_est169:*
250: gb_est170:*
251: gb_est171:*
252: gb_est172:*
253: gb_est173:*
254: gb_est174:*
255: gb_est175:*
256: gb_est176:*
257: gb_est177:*
258: gb_est178:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	18	75.0	426	243	AZ422893	AZ422893 IM0201G13
C 2	17	70.8	406	243	AM528260	AM528260 UT-R-BT1-
C 3	17	70.8	591	232	AQ075173	AQ075173 HS-5520_B
C 4	16	66.7	159	232	AA528604	AA528604 nF02C09..S
C 5	16	66.7	210	225	AQ237895	AQ237895 RRC111-68
C 6	16	66.7	266	103	AI914472	AI914472 w49e06.x
C 7	16	66.7	268	160	BB528140	BB528140 BB528140
C 8	16	66.7	289	134	BB473121	BB473121 BB473121
C 9	16	66.7	294	3	AA174030	AA174030 zp01906..T
C 10	16	66.7	295	2	AA077273	AA077273 7B11G05..C
C 11	16	66.7	337	15	AI079841	AI079841 ox50d02..x
C 12	16	66.7	427	107	AU040896	AU040896 AU040896
C 13	16	66.7	439	107	AU017418	AU017418 AU017418
C 14	16	66.7	451	107	AU045114	AU045114 AU045114
C 15	16	66.7	489	10	AA682402	AA682402 zj86d05..s
C 16	16	66.7	497	225	AQ225246	AQ225246 HS-2012..B
C 17	16	66.7	498	107	AU084305	AU084305 AU084305
C 18	16	66.7	506	228	AQ440938	AQ440938 HS-5098..B
C 19	16	66.7	511	230	AQ547803	AQ547803 RRC1-11-3
C 20	16	66.7	530	107	AU085116	AU085116 AU085116
C 21	16	66.7	571	107	AU045115	AU045115 AU045115
C 22	16	66.7	573	107	AU023320	AU023320 AU023320
C 23	16	66.7	593	230	AQ548670	AQ548670 RRC1-11-3
C 24	16	66.7	606	250	AZ856748	AZ856748 2M0161H18
C 25	16	66.7	621	239	AZ183973	AZ183973 SP-1002..A
C 26	16	66.7	658	237	AZ062577	AZ062577 RRC1-23-4
C 27	16	66.7	694	139	BE758383	BE758383 SWYACAL08
C 28	16	66.7	723	230	AZ856834	AZ856834 2M0161H20
C 29	16	66.7	740	259	AZ184562	AZ184562 SP-1003..A
C 30	16	66.7	789	166	BE305921	BE305921 601101687
C 31	16	66.7	791	218	AF010852	AF010852 AF010852
C 32	16	66.7	854	137	BE569058	BE569058 AF010852
C 33	16	66.7	1023	110	CNS01PE0	CNS01PE0 AL154696
C 34	16	66.7	141	210	AM032866	AM032866 EST276425
C 35	16	66.7	196	108	AU180439	AU180439 AU180439
C 36	15	62.5	202	37	AW109292	AW109292 g9t00003K
C 37	15	62.5	205	162	BE054208	BE054208 GA_Ea000
C 38	15	62.5	206	246	AZ609907	AZ609907 IM0434122
C 39	15	62.5	212	153	BG442137	BG442137 GA_Ea001
C 40	15	62.5	227	161	BB566737	BB566737 BB566737
C 41	15	62.5	229	103	AI915011	AI915011 t997b12..x
C 42	15	62.5	240	126	BB149447	BB149447 BB149447
C 43	15	62.5	260	157	D79868	D79868 H0M345A06B
C 44	15	62.5	264	132	BB354670	BB354670 BB354670
C 45	15	62.5	287	161	BB592792	BB592792 BB592792

ALIGNMENTS

RESULT 1
LOCUS AZ422893/c
DEFINITION 1M0201G13R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
ACCESSION AZ422893
VERSION AZ422893.1 GI:10546906
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 426)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duvall,B., Hamli,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Petersen,T., Kelly,
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb
JOURNAL Plasmid inserts
COMMENT Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0201 row: G column: 13
Seq primer: CACACGACCAACACGATGACC
Class: plasmid ends
High quality sequence stop: 426.

FEATURES
source
1..426
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6j"
/db_xref="taxon:10090"
/clone="UUGC1M0201G13"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/anases/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PMD42 (g14732149b1aP129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 130 a 78 c 95 g 123 t
ORIGIN

Query Match 75.0%; Score 18; DB 243; Length 426;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ttatlttggtcact 23
Db 164 TTTATTTTGGCTTCAC 147

RESULT 2
LOCUS AM528260
DEFINITION AM528260 406 bp mRNA EST
UT-R-BT1-ajw-e-04-0-UI-s1 UT-R-BT1 Rattus norvegicus cDNA clone
ACCESSION AM528260
VERSION AM528260.1 GI:7170674
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 406)
Bonaldi,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene

JOURNAL
MEDLINE
COMMENT

discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NCI site and the oligo-dT track served to identify it as a clone from the normalized corpus striatum library cDNA library preparation: M.B. Soares Lab Clone Distribution: clones will be available through Research Genetics (www.resgen.com) The following repetitive elements were found in this cDNA sequence: 1-29,
>AT_rich_low_complexity 64-178, >KSIINE#SINE/B4(B5)
Seq primer: M13 Forward
POLYA=yes.

FEATURES

Location/Qualifiers

1..406
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BT1-ajw-e-04-0-UI"
/clone_lib="UI-R-BT1"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pYT3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The library UI-R-BT1 is a subtracted library derived from a mixture of the following tissues: hippocampus, thalamus, mid-brain, medulla, corpus striatum, cerebral cortex and testis. For a detailed description of the library from which this clone was derived, please visit our web site at ratseq.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG_LIB=UI-R-BT1
TAG_TISSUE=corpus-striatum
TAG_SEQ=CTAGC
BASE COUNT 96 a 113 c 82 g 115 t
ORIGIN

Query Match 70.8%; Score 17; DB 117; Length 406;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 agattatttggcctt 19
|||||
Db 19 AGATTATTGGGCTT 35

RESULT 3
LOCUS AO705173 591 bp DNA GSS 07-JUL-1999
DEFINITION HS.5520.B1.E02.F7A.RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate-1096 Col-3 Row-J, DNA sequence.
ACCESSION AO705173
VERSION AO705173.1 GI:5414599
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 591)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

TITLE
JOURNAL
MEDLINE
COMMENT

Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Research Genetics (info@resgen.com). BAC end Web Server: http://www.hsc.washington.edu/plate:1096_row:J_column:3
Seq primer: T7
Class: BAC ends
High quality sequence stop: 591.

FEATURES

Location/Qualifiers

1..591
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate-1096 Col-3 Row-J"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites"
BASE COUNT 128 a 137 c 103 g 216 t 7 others
ORIGIN

Query Match 70.8%; Score 17; DB 232; Length 591;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 tatttggcttacc 24
|||||
Db 281 TATTTGGCTTCACTC 297

RESULT 4
LOCUS AA528604/c 159 bp mRNA EST 20-AUG-1997
DEFINITION nt02c09.s1 NCI-CGAP_K1d1 Homo sapiens cDNA clone IMAGE:912592, mRNA sequence.
ACCESSION AA528604
VERSION AA528604.1 GI:2270673
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 159)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Kitzman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bdrp/image/image.html

JOURNAL
COMMENT

Insert Length: 270 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham.

FEATURES

Location/Qualifiers
source

1. 159
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:912592"
/clone_1lb="NCI CGAP Kid1"
/tissue_type="Kidney"
/lab_host="DH10B"
/note="Vector: PAMP10; mRNA made from invasive kidney tumor, cDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. Reference: Krizman et al. (1996) Cancer Research 56:5380-5383."
BASE COUNT 53 a 27 c 23 g 56 t
ORIGIN

Query Match 66.7%; Score 16; DB 8; Length 159;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 gattatttggcctt 19
|||||
Db 77 GATTATTGCGCTT 62

RESULT 5
LOCUS AQ237895/c 210 bp DNA GSS 21-APR-1999
DEFINITION RPII11-66J13. TK RPII-11 Homo sapiens genomic clone RPII-11-68J13,
DNA sequence.
VERSION AQ237895.1 GI:3670186
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 210)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Baas,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
TITLE Use of human BAC End Sequences for Sequence-Ready Map Building
JOURNAL Unpublished (1998)
COMMENT Other-GSSs: RPII11-68J13.TJ
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org

Clones are derived from the human BAC library RPII-11. For BAC
library availability, please contact Pieter de Jong
BACPAC Resources (http://bacpac.med.bufile.edu). Clones may be purchased from
Research Genetics (http://inforesgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: TJ
Class: BAC ends

FEATURES

Location/Qualifiers
source

1. 210
/organism="Homo sapiens"
/db_xref="GDB:7525956"
/db_xref="taxon:9606"
/clone="RPII-11-68J13"
/clone_1lb="RPII-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: PBACE3.6; Site.1: EcoRI; Site.2: EcoRI;
RPII11 Human Male BAC Library"
BASE COUNT 85 a 20 c 32 g 71 t 2 others

ORIGIN

Query Match 66.7%; Score 16; DB 225; Length 210;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 attatttggcctc 20
|||||
Db 142 ATTATTGCGCTTC 127

RESULT 6
LOCUS A1914472 266 bp mRNA EST 17-DEC-1999
DEFINITION w49e06.x1 Soares NFL.T GBC.S1 Homo sapiens cDNA clone
IMAGE:2331490.3' similar to gb:X63547.cd82 UBROUTIN
CARBOXYL-TERMINAL HYDROLASE TRE-2 (HUMAN);, mRNA sequence.
A1914472
A1914472.1 GI:5634327
VERSION EST.
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 266)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 421 Std Error: 0.00
Seq primer: -400P from GIbco.

FEATURES

Location/Qualifiers
source

1. 266
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:2331490"
/clone_1lb="Soares_NFL.T_GBC.S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site.1: Not I; Site.2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCI-CGAP-GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The diver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo.
BASE COUNT 61 a 57 c 45 g 103 t
ORIGIN

Query Match 66.7%; Score 16; DB 103; Length 266;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ttatttggcctca 21
|||||
Db 43 TTATTGCGCTTCA 58

RESULT 7
LOCUS BB528140 268 bp mRNA EST 28-JUL-2000
DEFINITION BB528140 RIKEN full-length enriched, 15 days embryo head Mus
musculus cDNA clone D930045A20.3', mRNA sequence.

ACCESSION BB528140
 VERSION BB528140.1 GI:9579598
 KEYWORDS
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 268)
 Kono, H., Akawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Horii, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya, T., Tsunoda, Y., Watabiki, A., Watanabe, S., Yamamura, T., Yamanaka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.
 RIKEN Mouse ESTs (Kono, H., et al.)
 Unpublished (2000)
 Contact: Yoshihide Hayashizaki
 Genome Exploration Research Group, Life Science Tsukuba Center,
 The Institute of Physical and Chemical Research (RIKEN), Genomic
 Sciences Center
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
 Tel: +81-298-36-9013
 Fax: +81-298-36-9098
 Email: genome-res@rtc.riken.go.jp/
 URL: http://genome.rtc.riken.go.jp/
 Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
 Thermotranscription and thermotranscription of thermolabile enzymes by
 thermotranscription and its application for the synthesis of full length
 cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
 Itoh, M., Katsunari, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
 Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y., and Hayashizaki, Y.
 Automated filtration-based high-throughput plasmid preparation
 system. Genome Res. 9 (5), 463-470 (1999)
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
 19-44 (1999)
 Please visit our web site (<http://genome.rtc.riken.go.jp>) for
 further details.

FEATURES

source

1. 268
 location/Qualifiers
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="D930045A20"
 /clone_1lb="RIKEN full-length enriched, 15 days embryo
 head"
 /sex="mixed"
 /tissue="type="head"
 /dev_stage="15 days embryo"
 /lab_host="DH10B"
 /note="Site 1: Salt; Site 2: BamHI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN. Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer 15'
 GAGAGAGAGATTCGAGTTATTAATTAATCCCCCCCCCCC 3'. cDNA
 was cloned into the XhoI and BamHI sites. Vector: a

modified plasmid KS(+) after bulk excision from Lambda
 FLX 1"

Query Match 66.7%; Score 16; DB 160; Length 268;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ttatttgggttcac 22
 Db 120 TTTATTTGGGCTTCAC 135

RESULT 8
 BB473121 289 bp mRNA EST 22-JUL-2000
 BB473121 RIKEN full-length enriched, 12 days embryo eyeball Mus
 musculus cDNA clone D230049L08 3', mRNA sequence.

ACCESSION BB473121 GI:9390310
 VERSION BB473121
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 289)
 Kono, H., Akawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Horii, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya, T., Tsunoda, Y., Watabiki, A., Watanabe, S., Yamamura, T., Yamanaka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.
 RIKEN Mouse ESTs (Kono, H., et al.)
 Unpublished (2000)
 Contact: Yoshihide Hayashizaki
 Genome Exploration Research Group, Life Science Tsukuba Center,
 The Institute of Physical and Chemical Research (RIKEN), Genomic
 Sciences Center
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
 Tel: +81-298-36-9013
 Fax: +81-298-36-9098
 Email: genome-res@rtc.riken.go.jp/
 URL: http://genome.rtc.riken.go.jp/
 Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
 Thermotranscription and thermotranscription of thermolabile enzymes by
 thermotranscription and its application for the synthesis of full length
 cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
 Itoh, M., Katsunari, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
 Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y., and Hayashizaki, Y.
 Automated filtration-based high-throughput plasmid preparation
 system. Genome Res. 9 (5), 463-470 (1999)
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
 19-44 (1999)
 Please visit our web site (<http://genome.rtc.riken.go.jp>) for
 further details.

FEATURES

source

1. 289
 location/Qualifiers
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="D230049L08"
 /clone_1lb="RIKEN full-length enriched, 12 days embryo
 head"
 /sex="mixed"
 /tissue="type="head"
 /dev_stage="15 days embryo"
 /lab_host="DH10B"
 /note="Site 1: Salt; Site 2: BamHI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN. Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer 15'
 GAGAGAGAGATTCGAGTTATTAATTAATCCCCCCCCCCC 3'. cDNA
 was cloned into the XhoI and BamHI sites. Vector: a

```

eyeball"
/tissue_type="eyeball"
/dev_stage="12 days embryo"
/lab_host="DH10B"
/Note="Site 1: Sali; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGCGCGCGACACGAGATTTTCTTTTCTTTTCTTTT 3']. cDNA was
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGAGATTCGAGTATTAATTAATCCCGCCCCCCCC 3']. cDNA
was cleaved with BamHI and XhoI. Vector: a modified
pBluescript KS(+) after bulk excision from Lambda FIC I."
BASE COUNT      58 a      70 c      60 g      101 t
ORIGIN

Query Match      66.7%; Score 16; DB 134; Length 289;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 7 ttatttggcttcac 22
|||||
Db 142 TTATTTTGGCCTTCA 157

RESULT 9
AA174030      294 bp      mRNA      EST      09-MAR-1998
LOCUS      zpo1906.r1 Stratigene ovarian cancer (#937219) Homo sapiens cDNA
DEFINITION      clone IMAGE:595162 5', mRNA sequence.
ACCESSION      AA174030
VERSION      AA174030.1 GI:1754180
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 294)
AUTHORS      Hillier, L., Allen, M., Bowles, L., Dubnue, T., Geisler, G., Jost, S.,
Krizman, D., Kucaba, M., Lacy, M., Le, N., Lennon, G., Maier, M., Martin
White, Y., Wylie, T., Waterston, R. and Wilson, R.
WashU-NCI human EST Project
Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1420 Std Error: 0.00
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 260.
Location/Qualifiers
1..294
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:595162"
/clone_1lb="Stratigene ovarian cancer (#937219)"
/sex="female"
/dev_stage="adult, 64 years"
/lab_host="SOLR (kanamycin resistant)"
/Note="Vector: Bluescript SK-; Site 1: EcoRI; Site 2: XhoI
; Cloned unidirectionally. Primer: Oligo dT. Papillary

```

```

serous carcinoma, isolated from ascites, 64 year old
caucasian. Average insert size: 0.8 kb; Uni-ZAP XR Vector;
-5' adaptor sequence: 5' GATTTCGACAGAG 3' -3' adaptor
sequence: 5' CTCGAGCTTTTCTTTTCTTTTCTTTT 3"
BASE COUNT      107 a      50 c      41 g      95 t      1 others
ORIGIN

Query Match      66.7%; Score 16; DB 3; Length 294;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 ttatttggcttca 21
|||||
Db 265 TTATTTTGGCCTTCA 280

RESULT 10
AA077273      295 bp      mRNA      EST      24-SEP-1999
LOCUS      7B11G05 Chromosome 7 Fetal Brain cDNA Library Homo sapiens cDNA
DEFINITION      clone 7B11G05, mRNA sequence.
ACCESSION      AA077273
VERSION      AA077273.1 GI:1836747
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 295)
AUTHORS      Touchman, J.W., Bouffard, G.G., Weintrub, L.A., Idol, J.R., Wang, L.,
Robbins, C.M., Nussbaum, J.C., Lovett, M. and Green, E.D.
2006 expressed-sequence tags derived from human chromosome
7-enriched cDNA libraries
Genome Res. 7 (3), 281-292 (1997)
97228905
CONTACT: Eric D. Green
Genome Technology Branch
National Human Genome Research Institute/NIH
49 Convent Dr., MSC4431, Building 49, Room 2A08, Bethesda, MD 20892
Tel: 3014020201
Fax: 3014024735
Email: egreen@nhgri.nih.gov
Plate: 11 row: G column: 05
Seq primer: -21M13 (ABI).
Location/Qualifiers
1..295
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="7B11G05"
/clone_1lb="Chromosome 7 Fetal Brain cDNA Library"
/sex="female and male mixture"
/tissue_type="Brain"
/dev_stage="pool of 9 week and 12 week"
/lab_host="E. coli strain DH5 alpha"
/Note="Organ: brain; Vector: pMP10; cDNA was generated
from cytoplasmic RNA using a mixture of random DNA
hexamers and oligo(dT). From this pool of cDNA, human
chromosome 7-enriched cDNA was isolated by direct cDNA
selection using chromosome 7 genomic DNA (cosmids). The
resulting direct-selected cDNA was cloned into a plasmid
vector using a non-directional uracil DNA glycosylase (UDG)
-mediated cloning strategy."
BASE COUNT      63 a      56 c      76 g      93 t      7 others
ORIGIN

Query Match      66.7%; Score 16; DB 2; Length 295;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4 gattatttggctt 19
|||||

```

Db 166 GATTATATTTGGGCTT 181

RESULT 11
LOCUS A1079841
DEFINITION ox50d02.x1 Soares_total_fetus_Nb2HF8_9w EST 28-AUG-1998
IMAGE:1659747 3', mRNA sequence.
ACCESSION A1079841
VERSION A1079841.1 GI:3416092
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 337)
AUTHORS NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 896 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 281.
Location/Qualifiers
1..337
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1659747"
/clone_1lb="Soares_total_fetus_Nb2HF8_9w"
/dev_stage="8-9 weeks"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from pooled 8-9 week (total) fetus material with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGCGGCGCCCTTAATTTTCTTTTCTTTT 3']
Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Benito Soares and M. Fatima Bonaldo."

BASE COUNT 119 a 63 c 51 g 104 t

ORIGIN

Query Match 66.7%; Score 16; DB 15; Length 337;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 agattattttgggctt 18
|||||

Db 36 AGATTATATTTGGGCTT 51

RESULT 12
LOCUS AU040896/c
DEFINITION AU040896 427 bp mRNA EST 04-DEC-1998
AU040896 Mouse four-cell embryo cDNA Mus musculus cDNA clone
J0820C10 3', mRNA sequence.
ACCESSION AU040896
VERSION AU040896.1 GI:3954720
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 427)
AUTHORS Ko, M.S.H., Kitchen, J.R., Wang, X., Threat, T.A., Sun, T., DePalma, G.E., Liang, Y., Kargul, G.J., Sharata, R., Lim, M.K. and Doi, H.

TITLE
JOURNAL Systematic analyses of genes expressed in 4-cell mouse embryo (The ERATO/Doi Project at Wayne State University)
COMMENT Unpublished (1998)
Contact: Hirofumi Doi
Doi Biosymmetry Project, ERATO
Japan Science and Technology Corporation (JST)
WBG Marive East 12F, 2-6 Nakase, Mihama-ku, Chiba 261-71, Japan
Email: hdoe@doe.jst.go.jp

FEATURES
source
1..427
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="J0820C10"
/clone_1lb="Mouse four-cell embryo cDNA"
/dev_stage="four-cell-embryo"
Location/Qualifiers
123 a 62 c 107 g 135 t

BASE COUNT 123 a 62 c 107 g 135 t

ORIGIN

Query Match 66.7%; Score 16; DB 107; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 ttatttgggcttcac 22
|||||

Db 419 TTATTTTGGGCTTCAC 404

RESULT 13
LOCUS AU017418/c
DEFINITION AU017418 439 bp mRNA EST 19-OCT-1998
AU017418 Mouse two-cell stage embryo cDNA Mus musculus cDNA clone
J0738H04 3', mRNA sequence.
ACCESSION AU017418
VERSION AU017418.1 GI:3372908
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 439)
AUTHORS Ko, M.S.H., Kitchen, J.R., Wang, X., Threat, T.A., Sun, T., DePalma, G.E., Liang, Y., Kargul, G.J., Sharata, R. and Doi, H.
Systematic analyses of genes expressed in 2-cell stage mouse embryos (The ERATO/Doi Project at Wayne State University) (Ko, M.S.H. et al.)
JOURNAL Unpublished (1998)
COMMENT Contact: Hirofumi Doi
Doi Biosymmetry Project, ERATO
Japan Science and Technology Corporation (JST)
WBG Marive East 12F, 2-6 Nakase, Mihama-ku, Chiba 261-71, Japan
Email: hdoe@doe.jst.go.jp

FEATURES
source
1..439
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="J0738H04"
/clone_1lb="Mouse two-cell stage embryo cDNA"
/dev_stage="two-cell stage embryo"
Location/Qualifiers
128 a 62 c 109 g 138 t 2 others

BASE COUNT 128 a 62 c 109 g 138 t 2 others

ORIGIN

Query Match 66.7%; Score 16; DB 107; Length 439;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 ttatttgggcttcac 22
|||||

Db 420 TTATTTTGGGCTTCAC 405

us-09-663-020-7.rst

IMAGE Consortium (infoelimage.llnl.gov) for further information
Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 427.

1.489

/organism- "Homo sapiens"

```
/clone="IMAGE:461769"
```

```
soares_rectal_river_spleen_infls_sl
/sex="male"
```

```
/dev_stage="20 week-post conception fetus"
```

/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)

This is a subtracted version of the original Soares fetal

liver spleen infus library. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5'

double-stranded cDNA was ligated to Eco RI adapter

(Pharmacacia), digested with Pac I and cloned into the Pac I

went through one round of normalization. Library

constructed by Bento Soares and M. Fatima	114 t	1	Others
112 c	135 g		

100.0%;	Pred. No. 1.1e+02,
---------	--------------------

ervative 0; Mismatches 0; Indels 0; Gaps 0;

rectca 21
111111

SCITCA 25

Search completed: November 15, 2001, 06:39:15
Job time: 9291 sec

zj86d05.s1 Soares_fetal_liver_spleen_INTLS_S1 Homo sapiens CDNA

HISTOCOMPATIBILITY ANTIGEN, DR-1 BETA CHAIN (HUMAN);, mRNA sequence.

AA682402
AA682402.1 GI:2669683

EST.
human

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontari

1 (bases 1 to 489)
Miller L. Allen M. Bowles R. Debusse M. G. J. C.

Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, R., Moore, R., Schellenberg, R., Stoto, M., and T...

White, I., Wylie, T., Waterston, R. and Wilson, R.
WashU-NCI human EST Project

Unpublished (1997)
Contact: WILSON RK

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estevatson.wustl.edu
This clone is available royalty-free through LNL; com

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 15, 2001, 06:39:15 ; Search time 1315.37 Seconds
(without alignments)
129.356 Million cell updates/sec

```
Title: 'US-09-663-020-8
Perfect score: 18
Sequence: 1 ggcgcagcttcagtgc 18
```

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 10228115 seqs, 4726426750 residues

Word size : 0

Total number of hits satisfying chosen parameters: 20456230

```
Minimum DB seq length: 0
```

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

```
Database :
EST.*
1:  qb_est1:*
2:  qb_est2:*
3:  qb_est3:*
4:  qb_est4:*
5:  qb_est5:*
6:  qb_est6:*
7:  qb_est7:*
8:  qb_est8:*
9:  qb_est9:*
10: qb_est10:*
11: qb_est11:*
12: qb_est12:*
13: qb_est13:*
14: qb_est14:*
15: qb_est15:*
16: qb_est16:*
17: qb_est17:*
18: qb_est18:*
19: qb_est19:*
20: qb_est20:*
21: qb_est21:*
22: qb_est22:*
23: qb_est23:*
24: qb_est24:*
25: qb_est25:*
26: qb_est34:*
27: qb_est35:*
28: qb_est36:*
29: qb_est37:*
30: qb_est38:*
31: qb_est39:*
32: qb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthuni:*
36: em_esthuni2:*
37: em_esthuni3:*
38: em_esthuni4:*
39: em_esthuni5:*
40: em_esthuni6:*
41: em_esthuni7:*
42: em_esthuni8:*
43: em_esthuni9:*
```

```
117: gb_est148:*
118: gb_est149:*
119: gb_est150:*
120: gb_est151:*
121: gb_est152:*
122: gb_est153:*
123: gb_est154:*
124: gb_est155:*
125: gb_est156:*
126: gb_est157:*
127: gb_est158:*
128: gb_est159:*
129: gb_est160:*
130: gb_est161:*
131: gb_est162:*
132: gb_est163:*
133: gb_est164:*
134: gb_est165:*
135: gb_est166:*
136: gb_est167:*
137: gb_est168:*
138: gb_est169:*
139: gb_est170:*
140: gb_est171:*
141: gb_est172:*
142: gb_est173:*
143: gb_est174:*
144: gb_est175:*
145: gb_est176:*
146: gb_est177:*
147: gb_est178:*
148: gb_est179:*
149: gb_est180:*
150: gb_est181:*
151: gb_est182:*
152: gb_est183:*
153: gb_est184:*
154: gb_est185:*
155: gb_est186:*
156: gb_est187:*
157: gb_est188:*
158: gb_est189:*
159: gb_est190:*
160: gb_est191:*
161: gb_est192:*
162: gb_est193:*
163: gb_est194:*
164: gb_est195:*
165: gb_est196:*
166: gb_est197:*
167: gb_est198:*
168: gb_est199:*
169: gb_est200:*
170: gb_est201:*
171: gb_est202:*
172: gb_est203:*
173: gb_est204:*
174: gb_est205:*
175: gb_est206:*
176: gb_est207:*
177: gb_est208:*
178: gb_est209:*
179: gb_est210:*
180: gb_est211:*
181: gb_est212:*
182: gb_est213:*
183: gb_est214:*
184: gb_est215:*
185: gb_est216:*
186: gb_est217:*
187: gb_est218:*
188: gb_est219:*
189: gb_est220:*

190: gb_est110:*
191: gb_est111:*
192: gb_hic:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_inv4:*
207: em_gss_inv5:*
208: em_gss_inv6:*
209: em_gss_inv7:*
210: em_gss_inv8:*
211: em_gss_inv9:*
212: em_gss_inv10:*
213: em_gss_inv11:*
214: em_gss_inv12:*
215: em_gss_inv13:*
216: em_gss_inv14:*
217: em_gss_inv15:*
218: em_gss_inv16:*
219: em_gss_inv17:*
220: em_gss_inv18:*
221: em_gss_inv19:*
222: em_gss_inv20:*
223: em_gss_inv21:*
224: em_gss_inv22:*
225: em_gss_inv23:*
226: em_gss_inv24:*
227: em_gss_inv25:*
228: em_gss_inv26:*
229: em_gss_inv27:*
230: em_gss_inv28:*
231: em_gss_inv29:*
232: em_gss_inv30:*
233: em_gss_inv31:*
234: em_gss_inv32:*
235: em_gss_inv33:*
236: em_gss_inv34:*
237: em_gss_inv35:*
238: em_gss_inv36:*
239: em_gss_inv37:*
240: em_gss_inv38:*
241: em_gss_inv39:*
242: em_gss_inv40:*
243: em_gss_inv41:*
244: em_gss_inv42:*
245: em_gss_inv43:*
246: em_gss_inv44:*
247: em_gss_inv45:*
248: em_gss_inv46:*
249: em_gss_inv47:*
250: em_gss_inv48:*
251: em_gss_inv49:*
252: em_gss_inv50:*
253: em_gss_inv51:*
254: em_gss_inv52:*
255: em_gss_inv53:*
256: em_gss_inv54:*
257: em_gss_inv55:*
258: em_gss_inv56:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	18	100.0	67	4	AA251974	AA251974 z64601.s
C 2	18	100.0	72	110	AM007890	AM007890 wv46e10.x
C 3	18	100.0	75	159	N38861	N38861 yy80n11.s1
C 4	18	100.0	86	158	H39321	H39321 DR20 IFNgam
C 5	18	100.0	90	12	AA828120	AA828120 cd71a01.s
C 6	18	100.0	90	103	AI864984	AI864984 wk06c11.x
C 7	18	100.0	95	7	AA457423	AA457423 aa86b02.r
C 8	18	100.0	101	12	AA835205	AA835205 ak64n01.s
C 9	18	100.0	102	9	AA604960	AA604960 no93b10.s
C 10	18	100.0	103	4	AA228795	AA228795 nc14e07.s
C 11	18	100.0	105	4	AA247810	AA247810 hfe0642.s
C 12	18	100.0	105	145	BF194862	BF194862 7089e07.x
C 13	18	100.0	107	159	N23686	N23686 yw46a02.s1
C 14	18	100.0	107	159	N23691	N23691 yw46b01.s1
C 15	18	100.0	110	157	D82540	D82540 HUMHBC4284
C 16	18	100.0	111	139	N29962	N29962 yw53f06.s1
C 17	18	100.0	114	20	AI433008	AI433008 th42e09.x
C 18	18	100.0	118	8	AA484085	AA484085 ne73c10.s
C 19	18	100.0	118	158	H67529	H67529 yu68c11.s1
C 20	18	100.0	118	230	AO53772	AO53772 RPT-11-4
C 21	18	100.0	122	8	AA528383	AA528383 ne82n04.s
C 22	18	100.0	122	218	AF149534	AF149534 AF149534
C 23	18	100.0	126	19	AI357151	AI357151 qx62g10.x
C 24	18	100.0	126	118	AM574959	AM574959 ut-HF-BNO
C 25	18	100.0	128	102	AI824649	AI824649 wc48a05.x
C 26	18	100.0	133	23	AI696442	AI696442 tw61e03.x
C 27	18	100.0	135	9	AA573562	AA573562 nm52d10.s
C 28	18	100.0	136	12	BF983677	BF983677 602307244
C 29	18	100.0	138	278	AQ427220	AQ427220 CTBT-EL-
C 30	18	100.0	140	7	AA455670	AA455670 aa22f08.s
C 31	18	100.0	140	11	AA730009	AA730009 n289g05.s
C 32	18	100.0	141	1	AA017501	AA017501 ze38h02.r
C 33	18	100.0	142	2	AA091710	AA091710 m1216.seq
C 34	18	100.0	142	12	AA847496	AA847496 oe19a05.s
C 35	18	100.0	142	111	AM082537	AM082537 xb52e08.x
C 36	18	100.0	143	10	AA632600	AA632600 np82c09.s
C 37	18	100.0	143	18	AI305894	AI305894 qw73e10.x
C 38	18	100.0	144	18	AA668816	AA668816 ad41d11.s
C 39	18	100.0	144	18	AI306191	AI306191 qw74f02.x
C 40	18	100.0	144	18	AI306208	AI306208 qw74g10.x
C 41	18	100.0	145	257	B95154	B95154 CIT-HSP-217
C 42	18	100.0	146	12	AA826076	AA826076 oe24a11.s
C 43	18	100.0	148	120	AM769572	AM769572 wl66g04.x
C 44	18	100.0	149	103	AI922803	AI922803 ho14b04.x
C 45	18	100.0	151	6	AA346355	AA346355 ESM52550

ALIGNMENTS

RESULT 1
LOCUS AA251974/c
DEFINITION AA251974 67 bp mRNA EST 12-MAR-1997
z64601.s1 Soares.NhMPu.S1 Homo sapiens CDNA clone IMAGE:668184 3'
similar to gb:D25272.1111 ALU CLASS A WARNING ENTRY 1111 (HUMAN);
mRNA sequence.
AA251974.1 GI:1887144
EST.
Homo sapiens
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 67)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman
M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Maitra, M., Paterson, J.,
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston

TITLE
JOURNAL
COMMENT

R., Williamson, A., Wohlmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu

FEATURES

This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.lml.gov) for further information.
Trace considered overall poor quality
Seq primer: 41m13 fwd. ET from Amersham
High quality sequence stop: 1.
Location/Qualifiers
1. 67
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:668184"
/tissue_type="Soares.NhMPu.S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pTZ19-Pac
(Pharmacia) with a modified polylinker; Site: 1: Not I;
Site: 2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2bNH, pregnant uterus
NbHPU, and fetal heart NBH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

BASE COUNT
ORIGIN

10 a 21 c 18 g 18 t

Query Match 100.0%; Score 18; DB 4; Length 67;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggcgcagcttcagtcagtcag 18
|||||
Db 66 GCCGCAGCTTCAGTCAGCAG 49

RESULT 2
LOCUS AM007890/c
DEFINITION AM007890 72 bp mRNA EST 09-MAR-2000
wv46e10.x1 NCI-CGAP_Gas4 Homo sapiens CDNA clone IMAGE:2532618 3'
similar to contains Alu repetitive element; mRNA sequence.
AM007890
AM007890.1 GI:5856668
EST.
Homo sapiens
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 72)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps@femail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LML at:

www.bio.lnl.gov/bhrp/image/image.html
Insert Length: 508 Std Error: 0.00
Seq primer: -400p from Gibco
High quality sequence stop: 56.
Location/Qualifiers

FEATURES
source

1. 72
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="2532618"
/clone_lib="NCI-CCAP_Gas4"
/tissue_type="poorly differentiated adenocarcinoma with
signet ring cell features"
/lab_host="DH10B"
/note="Organ: Stomach; Vector: pCMV-SPORT6; Site: 1: SalI;
Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.69 kb. Life Technologies catalog #: 11549-011"

BASE COUNT 8 a 25 c 22 g 17 t
ORIGIN

Query Match 100.0%; Score 18; DB 110; Length 72;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggcggagcttcagtgag 18
|||||
Db 49 GCGCGAGCTTCAGTGAG 32

RESULT 3

LOCUS N38861 75 bp mRNA EST 19-JAN-1996
DEFINITION yy80h11.s1 Soares_multiple_sclerosis_2NBHMP Homo sapiens CDNA
clone IMAGE:279909 3' similar to contains Alu repetitive element;
mRNA sequence.
N38861
VERSION N38861.1 GI:1162068
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

ACCESSION

VERSION N38861.1 GI:1162068

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 75)
Haller, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman,
M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston,
R., Williamson, A., Wohlmann, P. and Wilson, R.
The Wash-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu

Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: m13 -40 forward
High quality sequence stop: 253.
Location/Qualifiers

FEATURES
source

1. 75

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="2532618"
/clone_lib="NCI-CCAP_Gas4"
/tissue_type="poorly differentiated adenocarcinoma with
signet ring cell features"
/lab_host="DH10B"
/note="Organ: Stomach; Vector: pCMV-SPORT6; Site: 1: SalI;
Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.69 kb. Life Technologies catalog #: 11549-011"

polylinker V-type: phagemid; Site: 1: Not I; Site: 2: Eco RI
1st strand cDNA was primed with a Not I - oligo(dT)
primer [5']
TGTACCATATCGAAGTGGAGCGCCGCAATTTTTTTTTTTTTTTT 3']
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot - 5. Library constructed by Bento
Soares and M. Fatima Bonaldo. RNA from 4 multiple sclerosis
lesions from one patient was kindly provided by Dr. Kevin
G. Becker (NINDS/NIH).

BASE COUNT 15 a 25 c 18 g 17 t
ORIGIN

Query Match 100.0%; Score 18; DB 159; Length 75;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggcggagcttcagtgag 18
|||||
Db 71 GCGCGAGCTTCAGTGAG 54

RESULT 4

LOCUS H39321 86 bp mRNA EST 27-AUG-1996
DEFINITION DR20 IFNgamma-induced astrocyte ESTs Rattus norvegicus CDNA clone
DR20 similar to complement C1 inhibitor, mRNA sequence.
H39321
VERSION H39321.1 GI:915111
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

1 (bases 1 to 86)
Kuchinke, W., Hart, R.P. and Jonakait, G.M.
Identification of mRNAs regulated by interferon-gamma in cultured
rat astrocytes by PCR differential display
Neuroimmunomodulation 2, 347-355 (1995)
96437728
Contact: Hart RP
Rutgers University
Dept of Biol Sci, Rutgers University, 101 Warren St, Newark NJ
07102
Tel: 2016481008
Fax: 2016481007
Email: thart@andromeda.rutgers.edu
Sequence is similar to complement C1 inhibitor gene (M30688) and
contains alu element
subcloned into pCRscript
Seq primer: KS

FEATURES
source

1. 86
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone_image="DR20"
/note="cDNA from IFNgamma-induced astrocyte ESTs"
was screened by PCR differential display. Regulated cDNAs
were amplified, sequenced and confirmed by dot-blot and
Northern.

BASE COUNT 17 a 25 c 25 g 19 t
ORIGIN

Query Match 100.0%; Score 18; DB 158; Length 86;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 99c9gagcttcagtgag 18
 |||||||
 Db 69 GCGCGAGCTTCGACGTGAG 52

RESULT 5
 AA828120 90 bp mRNA EST 07-APR-1998
 LOCUS AA828120/c
 DEFINITION cd71a01.s1 NCI_CGAP_Ov2 Homo sapiens cDNA clone IMAGE:1373352
 similar to contains Alu repetitive element/contains element TAR1
 repetitive element// mRNA sequence.

ACCESSION
 VERSION AA828120
 KEYWORDS AA828120.1 GI:2900483
 EST

SOURCE
 ORGANISM human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 90)

REFERENCE
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: c9apbs-remail.nih.gov
 Tissue Procurement: Christopher A. Moskaluk, M.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: David B. Krizman, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/bbrp/image/image.html
 Insert Length: 801 Std Error: 0.00
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 84.

FEATURES
 source
 Location/Qualifiers
 1..90
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1373352"
 /clone_lib="NCI_CGAP_Ov2"
 /sex="female"
 /tissue_type="ovary"
 /lab_host="DH10B"
 /note="Vector: PAMPI0; mRNA made from invasive ovarian
 tumor, cDNA made by oligo-dt priming. Non-directionally
 cloned. Size-selected on agarose gel, average insert size
 600 bp. Reference: Krizman et al. (1996) Cancer Research
 56:5380-5383."

BASE COUNT 14 a 26 c 21 g 29 t
 ORIGIN

Query Match 100.0%; Score 18; DB 12; Length 90;
 Best Local Similarity 100.0%; Pred. No. 2.2;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 99c9gagcttcagtgag 18
 |||||||
 Db 83 GCGCGAGCTTCGACGTGAG 66

RESULT 6
 A1864984 90 bp mRNA EST 21-DEC-1999
 LOCUS A1864984/c
 DEFINITION wk66c11.x1 NCI_CGAP_Lym12 Homo sapiens cDNA clone IMAGE:2411540 3'
 similar to contains Alu repetitive element// mRNA sequence.

ACCESSION
 VERSION A1864984
 KEYWORDS A1864984.1 GI:5529091
 EST
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 90)

REFERENCE
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: c9apbs-remail.nih.gov
 Life Technologies catalog #: 11547-015.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/bbrp/image/image.html
 Insert Length: 195 Std Error: 0.00
 Seq primer: -40UP from Gibco.

FEATURES
 source
 Location/Qualifiers
 1..90
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2411540"
 /clone_lib="NCI_CGAP_Lym12"
 /tissue_type="lymphoma, follicular mixed small and large
 cell"
 /lab_host="DH10B"
 /note="Organ: lymph node; Vector: PCMV-SPORT6; Site: 1:
 SalI; Site: 2: NotI; Cloned unidirectionally. Primer:
 oligo dt. Average insert size 1.25 kb. Life Technologies
 catalog #: 11547-015."

BASE COUNT 13 a 29 c 27 g 21 t
 ORIGIN

Query Match 100.0%; Score 18; DB 103; Length 90;
 Best Local Similarity 100.0%; Pred. No. 2.2;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 99c9gagcttcagtgag 18
 |||||||
 Db 74 GCGCGAGCTTCGACGTGAG 57

RESULT 7
 AA457423 95 bp mRNA EST 06-JUN-1997
 LOCUS AA457423/c
 DEFINITION aa66b02.r1 Striatogene fetal retina 937202.Homo sapiens cDNA clone
 IMAGE:838155 5' similar to contains Alu repetitive element// mRNA
 sequence.

ACCESSION
 VERSION AA457423
 KEYWORDS AA457423.1 GI:2180143
 EST

SOURCE
 ORGANISM human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 95)

REFERENCE
 AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
 Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,
 Schellenberg, K., Steptoe, M., Tan, F., Theisling, B., White, Y., Wylie,
 T., Waterston, R., and Wilson, R.
 WashU-Merck EST Project 1997
 Unpublished (1997)
 Contact: Wilson RK
 Washington University School of Medicine
 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu

TITLE
 JOURNAL This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -28m13 rev1 ET from Amersham
 High quality sequence stop: 76.

```

FEATURES
source
    1..95
        location/Qualifiers
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:838155"
            /clone_lib="Stratagene fetal retina 937202"
            /sex="mixed"
            /lab_host="SOJR (kanamycin resistant)"
            /note="Vector: pBluescript SK-; Site_1: EcoRI; site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Pooled retinal tissue. Average insert size: 1.0 kb: Uni-ZAP XR Vector: -5' adaptor sequence: 5' GATTCGCGACGCG 3'-3' adaptor sequence: 5' CTCAGCTTTTTTTTTTTTTTTTTT 3.'"
BASE COUNT
ORIGIN
    14 a          39 c          25 g          17 t

```

Query Match	100.0%	Score 18	DB 7	Length 95
Best Local Similarity	100.0%	Pred. No. 2.2		
Matches 18: Conservative	0	Mismatches	0	Indels
				Gaps 0
Qy	1	ggcggagcttcacgtgag	18	
Db	36	gacggagcttcgacgtgag	19	

RESULT 8
 AA835205/c
 LOCUS
 DEFINITION
 AA835205 101 bp mRNA EST 23-FEB-1998
 ak64h01.s1 Barsted pancreas HPRB1 Homo sapiens cDNA clone
 IMAGE:412689.3' similar to contains Alu repetitive element
 ; contains element KBR repetitive element ;, mRNA sequence.
 ACCESSION
 AA835205
 VERSION
 AA835205.1 GI:2208933
 KEYWORDS
 EST.
 SOURCE
 human.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota: Metazoa: Chordata: Craniata: Vertebrata; Euteleostomi: Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo. 1 (bases 1 to 101)
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisels, G., Jost, S., Kitzman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Mairr, M., Mart, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theisling, B., White, Y., Wyllie, T., Waterston, R., and Wilson, R.
WashU-NCI human EST Project
Unpublished (1997)
Contact: Wilson RK

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewartson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAB Consortium (info@image.llnl.gov) for further information.
Seq primer: -40m3 fwd. Et from Amersham.
Location/Qualifiers
1..101
source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1412689"
/clone_id="Barstead pancreas HPLMB1"
/sex="female"
/dev_stage="adult, 34 years"
/lab_host="DH10B"
/note="Organ: pancreas; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Site.1: EcoRI; Site.2: NotI;
1st strand cDNA was primed with a Not I - oligo(dT) primer
15',
TGTTCAGAACTCTCAACTGGAGCGCCGCCCCCTTTTTTTTTTTTTTTTTTTTTT
3.1; double-stranded cDNA was ligated to Eco RI adaptors
(AATTCGATCCCTTG), digested with Not I and cloned into the
Not I and Eco RI sites of the modified pT73 vector.
Library constructed by Bob Barstead."

```

BASE COUNT	14	a	36	c	27	g	24	t
ORIGIN								
Query Match	100.0%; Score 18; DB 12; Length 101;							
Best Local Similarity	100.0%; Pred. No. 2.2;							
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;								
Oy	1	ggcggagcttgcagftgag	18					
Db	72	GGCGGAGGCTTGCAGTGAG	55					

RESULT	9
AA604960/c	
LOCUS	AA604960
DEFINITION	nc93bd10..s1 NCI CGAP_P-2 Homo sapiens cDNA clone IMAGE:1114363
	similar to contains Alu repetitive element;contains element MER22
ACCESSION	AA604960
VERSION	AA604960.1
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
I (bases 1 to 102)
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.

Email: cgapds-remail.nih.gov
 Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuagui, M.D.,
 Michael Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: David B. Krizman, Ph.D.
 cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution Information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www.dio.llnl.gov/dbproj/image/image.html
 Insert length: 757 Std Error: 0.00
 Seq primer: ~40ml3 fwd. ET from Amersham
 High quality sequence stop: 88.
 Location/Qualifiers
 .1..102

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone IMAGE:1114363"
/clone_lib="NCI_COAP_Pir2"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/notes="Vector: PAMp10; Site_1: NotI; Site_2: EcoRI; 1st
strand cDNA was primed with oligo(dT)17 on 50 ng of
DNase-treated, total cellular RNA obtained from 5,000-10
,000 microdissected preneoplastic cells
histologically-determined to be prostatic intraepithelial
neoplasia 2 (PIN2) cells. Double-stranded cDNA was
ligated to EcoRI adaptors, 5 cycles of PCR applied to the
cDNA with an adaptor-specific primer, and the resulting
PCR product subcloned into PAMp10 by the UDG-cloning
method (Life Technologies). Average insert size is 600
bp. NOTE: Not directionally cloned. This library was
constructed by David Kitzman."

```

Query Match	100.0%	Score 18;	DB 9;	Length 102;
Best Local Similarity	100.0%;	Pred. No. 2.2;		
Matches 18;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 1 99c9gagcttcagtgag 18
 |||||||
 Db 82 GCGGAGCTTCAGTGAG 65

RESULT 10
 AA228795 103 bp mRNA EST 20-AUG-1997
 LOCUS AA228795/c
 DEFINITION nc14e07.s1 NCI_CGAP_Prl Homo sapiens cDNA clone IMAGE:1008132
 similar to contains Alu repetitive element/contains element MER28
 repetitive element // mRNA sequence.

ACCESSION AA228795
 VERSION AA228795.1 GI:1851455
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 103)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuang, M.D.
 Michael Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: David B. Krizman, Ph.D.
 cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/dbp/image/image.html
 Seq primer: -41m13 fwd. ET from Amersham
 High quality sequence stop: 81.
 Location/Qualifiers
 1..103
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="1008132"
 /clone_lib="NCI_CGAP_Prl"
 /sex="Male"
 /dev_stage="45 years old"
 /lab_host="DH10B"
 /note="Vector: PAMp10; Site_1: NotI; Site_2: EcoRI; 1st
 strand cDNA was primed with oligo(dT)17 on 50 ng of
 DNase-treated, total cellular RNA obtained from 5,000-10,
 000 microdissected, histologically normal prostate
 epithelial cells. Double-stranded cDNA was ligated to
 EcoRI adaptors, 5 cycles of PCR applied to the cDNA with
 an adaptor-specific primer, and the resulting PCR product
 subcloned into PAMp10 by the UDG-cloning method (Life
 Technologies). Average insert size is 600 bp. NOTE: Not
 directionally cloned. This library was constructed by
 David Krizman."

BASE COUNT 14 a 30 c 25 g 34 t
 ORIGIN

Query Match 100.0%; Score 18; DB 4; Length 103;
 Best Local Similarity 100.0%; Pred. NO. 2.2;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 99c9gagcttcagtgag 18
 |||||||
 Db 83 GCGGAGCTTCAGTGAG 66

RESULT 11
 AA247810 105 bp mRNA EST 11-MAR-1997
 LOCUS AA247810/c
 DEFINITION hfe0642.seq F Human fetal heart, Lambda ZAP Express Homo sapiens
 cDNA 5', mRNA sequence.

ACCESSION AA247810
 VERSION AA247810.1 GI:1880011
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 105)
 Liaw,C.C.
 CDNAS from human fetal heart (1997)
 JOURNAL Unpublished (1997)
 COMMENT Contact: Liew CC
 Department of Laboratory Medicine and Pathobiology
 University of Toronto
 Banting Institute, 100 College St., Toronto, Ontario, M5G1L5
 Tel: 4169788758
 Fax: 4169785650
 Email: liewcc@utcc.utoronto.ca
 PCR Primers
 FORWARD: 5' GCCAGCTCGAATTAACCCCTCAATAAGG 3'
 BACKWARD: 5' CCAGGATTTGTAATAGACACCTACATAAGGCG 3'
 Seq primer: 5' GAATTAACCCCTCAATAAGG 3'.
 Location/Qualifiers
 1..105
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="Human fetal heart, Lambda ZAP Express"
 /lab_host="E. coli XL1-blue"
 /note="Vector: Lambda ZAP Express; Site_1: EcoRI; Site_2:
 XhoI; mRNA was purified from human fetal hearts (8-10
 weeks). cDNA was synthesized using a XhoI-Oligo dT
 adaptor-primer. EcoRI adaptors were ligated, followed by
 digestion with XhoI, for directional cloning into
 predigested lambda ZAP Express."

BASE COUNT 20 a 40 c 26 g 19 t
 ORIGIN

Query Match 100.0%; Score 18; DB 4; Length 105;
 Best Local Similarity 100.0%; Pred. NO. 2.2;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 99c9gagcttcagtgag 18
 |||||||
 Db 28 GCGGAGCTTCAGTGAG 11

RESULT 12
 BF194862
 LOCUS BF194862
 DEFINITION 70899e07.x1 NCI_CGAP_KiD11 Homo sapiens cDNA clone IMAGE:3643500 3'
 similar to contains Alu repetitive element// mRNA sequence.
 ACCESSION BF194862
 VERSION BF194862.1 GI:11081053
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 105)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL, send email to:

FEATURES info@image.llnl.gov.
source Location/Qualifiers

1.105
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3643500"
/lab_host="NCI_CGAP_Kid11"
/lab_host="DH10B"

/note="Organ: Kidney; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP_Kid3 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 1322376-1323911, 1456007-1456775, and 1500552-1502855). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 29 a 19 c 31 g 26 t
ORIGIN

Query Match 100.0%; Score 18; DB 145; Length 105;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggcggagcttcagtcag 18
|||||
Db 46 GCGCGAGCTTCGAGTGC 63

RESULT 13
N23686 107 bp mRNA EST 28-DEC-1995
LOCUS N23686/c
DEFINITION yw46a02.s1 Weizmann Olfactory Epithelium Homo sapiens cDNA clone
IMAGE:255241 3' similar to contains Alu repetitive element; contains
element MERS repetitive element ;, mRNA sequence.

ACCESSION N23686
VERSION N23686.1 GI:1137836
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 107)
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins,
M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore,
B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,
Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevisan, E.,
Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)

TITLE JOURNAL
MEDLINE
COMMENT
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: m13 -40 forward
High quality sequence stop: 313.

FEATURES

source Location/Qualifiers
1.107
/organism="Homo sapiens"
/db_xref="GDB:366561"
/db_xref="taxon:9606"
/clone="IMAGE:255242"
/clone_id="Weizmann Olfactory Epithelium"
/sex="Female"

/tissue-type="Olfactory epithelium"
/dev_stage="35 year old"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: nose; Vector: pBluescript SK-; Site_1: EcoRI
; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo
dT. Olfactory epithelium, normal. Average insert size: 0.8
kb; Uni-ZAP XR Vector. Library constructed by N. Walker,
D. Lancel, Weizmann Institute of Science. -5' adaptor
sequence: 5' GAATTCGCGACGACG 3' -3' adaptor sequence: 5'
CCTCAGATTTTTTTTTTTTTTTT 3' "

BASE COUNT 15 a 42 c 24 g 23 t 3 others
ORIGIN

Query Match 100.0%; Score 18; DB 159; Length 107;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggcggagcttcagtcag 18
|||||
Db 72 GCGCGAGCTTCGAGTGC 55

RESULT 14
N23691 107 bp mRNA EST 28-DEC-1995
LOCUS N23691/c
DEFINITION yw46b01.s1 Weizmann Olfactory Epithelium Homo sapiens cDNA clone
IMAGE:255241 3' similar to contains Alu repetitive element; contains
element MERS repetitive element ;, mRNA sequence.

ACCESSION N23691
VERSION N23691.1 GI:1137841
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 107)
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins,
M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore,
B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,
Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevisan, E.,
Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)

TITLE JOURNAL
MEDLINE
COMMENT
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: m13 -40 forward
High quality sequence stop: 342.

FEATURES

source Location/Qualifiers
1.107
/organism="Homo sapiens"
/db_xref="GDB:366561"
/db_xref="taxon:9606"
/clone="IMAGE:255241"
/clone_id="Weizmann Olfactory Epithelium"
/sex="Female"
/tissue-type="Olfactory epithelium"
/dev_stage="35 year old"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: nose; Vector: pBluescript SK-; Site_1: EcoRI
; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo
dT. Olfactory epithelium, normal. Average insert size: 0.8
kb; Uni-ZAP XR Vector. Library constructed by N. Walker,
D. Lancel, Weizmann Institute of Science. -5' adaptor

sequence: 5' GAATTCGCCAGCAG 3' -3' adaptor sequence: 5'
 BASE COUNT 16 a 41 c 24 g 22 t 4 others
 ORIGIN

Query Match 100.0%; Score 18; DB 159; Length 107;
 Best Local Similarity 100.0%; Pred. No. 2.2;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgagcttcagctgag 18
 |||
 Db 70 GCGGAGCTTCAGCTGAG 53

RESULT 15

D82540 110 bp mRNA EST 09-FEB-1996
 LOCUS HDHBC4284 Human pancreatic Islet Homo sapiens cDNA, mRNA sequence.
 DEFINITION
 ACCESSION D82540
 VERSION D82540.1 GI:1183498
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 110)

AUTHORS Takeda,J.
 TITLE Human pancreatic islet ESTs
 JOURNAL Unpublished (1995)
 COMMENT Institute for Molecular and Cellular Regulation, Gunma University
 3-39-15 Showa-machi, Maebashi Gunma 371, Japan
 Tel: 272-20-8836
 Fax: 272-20-8896
 Email: jtakeda@b.gunma-u.ac.jp.

FEATURES
 source Location/Qualifiers
 1..110
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_id="Human pancreatic islet"
 /note="Vector: Lambda ZAPIT; Site.1: Eco RI; Site.2: Xho
 I; mRNA was prepared from normal adult human islets. cDNA
 was directionally synthesized from the Xho I in the vector
 to the EcoRI site. cDNA was size fractionated to remove
 sequences <1000 bp in size."

BASE COUNT 34 a 27 c 33 g 16 t
 ORIGIN

Query Match 100.0%; Score 18; DB 157; Length 110;
 Best Local Similarity 100.0%; Pred. No. 2.2;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgagcttcagctgag 18
 |||
 Db 29 GCGGAGCTTCAGCTGAG 46

Search completed: November 15, 2001, 06:39:19
 Job time: 9295 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 15, 2001, 06:17:26 ; Search time 1266.25 Seconds

(without alignments)
219.877 Million cell updates/sec

Title: US-09-663-020-8

Perfect score: 18
Sequence: 1 ggcggagcttcgactgag 18

Scoring table:

OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 7733874588 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba1:*
- 2: gb_ba2:*
- 3: gb_ba3:*
- 4: gb_in1:*
- 5: gb_in2:*
- 6: gb_in3:*
- 7: gb_om:*
- 8: gb_ov:*
- 9: gb_pat1:*
- 10: gb_pat2:*
- 11: gb_ph:*
- 12: gb_pl1:*
- 13: gb_pl2:*
- 14: gb_pl3:*
- 15: gb_pl4:*
- 16: em_ba1:*
- 17: em_ba2:*
- 18: em_fun:*
- 19: em_htgo_hum:*
- 20: em_htgo_inv:*
- 21: em_htgo_rod:*
- 22: em_htgo_hum1:*
- 23: em_htg_hum2:*
- 24: em_htg_hum3:*
- 25: em_htg_hum4:*
- 26: em_htg_hum5:*
- 27: em_htg_hum6:*
- 28: em_htg_hum7:*
- 29: em_htg_hum8:*
- 30: em_htg_inv1:*
- 31: em_htg_inv2:*
- 32: em_htg_other:*
- 33: em_htg_rod:*
- 34: em_hum1:*
- 35: em_hum2:*
- 36: em_hum3:*
- 37: em_hum4:*
- 38: em_hum5:*
- 39: em_hum6:*
- 40: em_hum7:*
- 41: em_in:*
- 42: em_om:*
- 43: em_or:*

- 44: em_ov:*
- 45: em_pat:*
- 46: em_ph:*
- 47: em_pl:*
- 48: em_ro:*
- 49: em_sts:*
- 50: em_sy:*
- 51: em_un:*
- 52: em_v1:*
- 53: gb_sts1:*
- 54: gb_sts2:*
- 55: gb_sts3:*
- 56: gb_sy:*
- 57: gb_un:*
- 58: gb_v11:*
- 59: gb_v12:*
- 60: gb_v12:*
- 61: gb_htg1:*
- 62: gb_htg2:*
- 63: gb_htg3:*
- 64: gb_htg4:*
- 65: gb_htg5:*
- 66: gb_htg6:*
- 67: gb_htg7:*
- 68: gb_htg8:*
- 69: gb_htg9:*
- 70: gb_htg10:*
- 71: gb_htg11:*
- 72: gb_htg12:*
- 73: gb_htg13:*
- 74: gb_htg14:*
- 75: gb_htg15:*
- 76: gb_htg16:*
- 77: gb_htg17:*
- 78: gb_htg18:*
- 79: gb_htg19:*
- 80: gb_htg20:*
- 81: gb_htg21:*
- 82: gb_htg22:*
- 83: gb_htg23:*
- 84: gb_htg24:*
- 85: gb_htg25:*
- 86: gb_pr1:*
- 87: gb_pr2:*
- 88: gb_pr3:*
- 89: gb_pr4:*
- 90: gb_pr5:*
- 91: gb_pr6:*
- 92: gb_pr7:*
- 93: gb_pr8:*
- 94: gb_pr9:*
- 95: gb_pr10:*
- 96: gb_pr11:*
- 97: gb_pr12:*
- 98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18	100.0	107	97	HUMALCE162
2	18	100.0	114	97	HUMALCE4
3	18	100.0	135	97	HUMALCE22
4	18	100.0	224	92	HS178H6F
5	18	100.0	234	93	HSU02063
6	18	100.0	235	54	HS248384
7	18	100.0	236	93	HSU02049
8	18	100.0	240	93	HSU02050

9 18 100.0 244 88 AF034649
10 18 100.0 245 93 HSIMDE1
11 18 100.0 248 89 AE270578
12 18 100.0 249 97 HUMALCE21
13 18 100.0 257 97 HS13B4F
14 18 100.0 262 54 G38317
15 18 100.0 262 97 HSU14706
16 18 100.0 262 97 HOMPON1115
17 18 100.0 263 97 PTU14717
18 18 100.0 267 97 S70689
19 18 100.0 267 97 S70692
20 18 100.0 277 92 HS5ALU475
21 18 100.0 278 97 S70697
22 18 100.0 279 97 S70694
23 18 100.0 281 97 S70699
24 18 100.0 286 92 HS21E1R
25 18 100.0 288 56 HSU14568
26 18 100.0 288 56 HSU14569
27 18 100.0 289 97 HSU14705
28 18 100.0 289 97 PTU14709
29 18 100.0 290 92 HS73F11R
30 18 100.0 292 97 HSU18388
31 18 100.0 294 93 HSALU012
32 18 100.0 299 93 HSALUCOL
33 18 100.0 300 97 HSU67818
34 18 100.0 301 97 HSU67818
35 18 100.0 302 92 HS47B2R
36 18 100.0 309 93 HSALU011
37 18 100.0 310 97 HSU18390
38 18 100.0 310 97 HSU67831
39 18 100.0 317 97 HSU67827
40 18 100.0 319 54 HS158A10T
41 18 100.0 320 10 HS15997
42 18 100.0 320 10 HS15997
43 18 100.0 320 97 HSU67831
44 18 100.0 321 93 HSU67831
45 18 100.0 321 93 HSU67831

ALIGNMENTS

AF034649 Tarsius s
Y08944 H.sapiens P
AE270578 Homo sapi
M87925 Human carc
Z60686 H.sapiens C
G38317 RCT-4-790D
U14706 Human Alu-S
U14717 Pan troglod
S70692 [repetitive
S70697 [repetitive
Z31000 H.sapiens D
S70697 [repetitive
S70694 [repetitive
S70699 [repetitive
Z60393 H.sapiens C
U14568 ***ALU WARR
U14569 ***ALU WARR
U14705 Human Alu-S
U14709 Pan troglod
Z62833 H.sapiens C
U18388 Human Alu s
X55933 Human DNA f
X75335 H.sapiens A
I47228 Homo sapien
U67818 Human prima
Z61176 H.sapiens C
X55932 Human DNA f
U18390 Human Alu s
U67831 Human prima
U16293 H.sapiens
I51997 Sequence 5
I72205 Sequence 5
D31746 Homo sapien
X54179 Human speci
Z77925 H.sapiens f

RESULT 1
LOCUS HUMALCE162 107 bp ss-RNA PRI 15-APR-1994
DEFINITION Human carcinoma cell-derived Alu RNA transcript, clone CE162.
ACCESSION M87924.1 GI:174871
VERSION M87924.1 GI:174871
KEYWORDS Alu repeat.
SOURCE Homo sapiens male embryo carcinoma cDNA to other RNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 107)
AUTHORS Sinnett,D., Richer,C., Deragon,J.-M. and Labuda,D.
TITLE Alu RNA transcripts in human embryonal carcinoma cells. Model of
JOURNAL post-transcriptional selection of master sequences
J. Mol. Biol. (1992) in press
FEATURES
SOURCE location/Qualifiers
1..107
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="Ntera2D1"
/dev_stage="embryo"
/sex="male"
/tissue_type="carcinoma"
BASE COUNT 28 a 30 c 35 g 14 t
ORIGIN

Query Match 100.0%; Score 18; DB 97; Length 107;
Best Local Similarity 100.0%; Pred. No. 0.79;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcggagcttcagtgag 18
|||||
Db 31 GCGGAGCTTCAGTGAG 48

RESULT 2
LOCUS HUMALCE44 114 bp ss-RNA PRI 15-APR-1994
DEFINITION Human carcinoma cell-derived Alu RNA transcript, clone CE44.
ACCESSION M87927.1 GI:174877
VERSION M87927.1 GI:174877
KEYWORDS Alu repeat.
SOURCE Homo sapiens male embryo carcinoma cDNA to other RNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 114)
AUTHORS Sinnett,D., Richer,C., Deragon,J.-M. and Labuda,D.
TITLE Alu RNA transcripts in human embryonal carcinoma cells. Model of
JOURNAL post-transcriptional selection of master sequences
J. Mol. Biol. (1992) in press
FEATURES
SOURCE location/Qualifiers
1..114
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="Ntera2D1"
/dev_stage="embryo"
/sex="male"
/tissue_type="carcinoma"
BASE COUNT 28 a 31 c 39 g 16 t
ORIGIN

Query Match 100.0%; Score 18; DB 97; Length 114;
Best Local Similarity 100.0%; Pred. No. 0.79;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 3
LOCUS HUMALCE22 135 bp ss-RNA PRI 15-APR-1994
DEFINITION Human carcinoma cell-derived Alu RNA transcript, clone CE22.
ACCESSION M87926.1 GI:174873
VERSION M87926.1 GI:174873
KEYWORDS Alu repeat.
SOURCE Homo sapiens male embryo carcinoma cDNA to other RNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 135)
AUTHORS Sinnett,D., Richer,C., Deragon,J.-M. and Labuda,D.
TITLE Alu RNA transcripts in human embryonal carcinoma cells. Model of
JOURNAL post-transcriptional selection of master sequences
J. Mol. Biol. (1992) in press
FEATURES
SOURCE location/Qualifiers
1..135
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="Ntera2D1"
/dev_stage="embryo"
/sex="male"
/tissue_type="carcinoma"
BASE COUNT 32 a 39 c 44 g 20 t
ORIGIN

Query Match 100.0%; Score 18; DB 97; Length 135;
Best Local Similarity 100.0%; Pred. No. 0.77;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcggagcttcagctgag 18
 |||
 Db 59 GCGCGAGCTTCAGCTGAG 76

RESULT 4
 H5178H6F 224 bp DNA PRI 19-OCT-1995
 LOCUS H.sapiens Cpg Island DNA genomic MseI fragment, clone 178h6,
 DEFINITION forward read cpj178h6.ftla.
 ACCESSION 259870
 VERSION 259870.1 GI:1031783
 KEYWORDS Cpg Island; genomic MseI fragment.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 224)
 AUTHORS Dodsorth,S.J., Huckle,E., Wilkinson,P. and Micklem,G.
 TITLE Direct Submission
 JOURNAL Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1RQ, England. E-mail contact: humquerry@sanger.ac.uk
 REFERENCE 2 (bases 1 to 224)
 AUTHORS Cross,S.H., Charlton,J.A., Nan,X. and Bird,A.P.
 TITLE Purification of Cpg Islands using a methylated DNA binding column
 JOURNAL Nat. Genet. 6 (3), 236-244 (1994)
 MEDLINE 94282070
 COMMENT Vector: pGEM-5zf(-)
 Clones are available from the UK MRC Human Genome Mapping Project
 Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:
 http://www.hgmp.mrc.ac.uk/ for details
 or contact: biolhelp@hgmp.mrc.ac.uk.
 FEATURES
 source
 1. 224
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /sex="male"
 /dev_stage="adult"
 /tissue_type="blood"
 /clone_lib="CGI-1"
 /clone="178h6"

BASE COUNT 56 a 61 c 55 g 50 t 2 others

ORIGIN

Query Match 100.0%; Score 18; DB 92; Length 224;
 Best Local Similarity 100.0%; Pred. No. 0.73;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcggagcttcagctgag 18
 |||
 Db 198 GCGCGAGCTTCAGCTGAG 181

RESULT 5
 HS002063 234 bp mRNA PRI 04-FEB-1999
 LOCUS Human clone 4 Alu repeat mRNA sequence.
 DEFINITION Human clone 4 Alu repeat mRNA sequence.
 ACCESSION U02063
 VERSION U02063.1 GI:406898
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 234)
 AUTHORS Liu,W.M., Marala,R.J., Rubin,C.M. and Schmid,C.M.
 TITLE Alu transcripts: cytoplasmic localisation and regulation by DNA
 methylation
 JOURNAL Nucleic Acids Res. 22 (6), 1087-1095 (1994)
 MEDLINE 94203794

REFERENCE 2 (bases 1 to 234)
 AUTHORS Liu,W.M.
 TITLE Direct Submission
 JOURNAL Submitted (23-SEP-1993) University of California, Davis, Molecular
 and Cellular Biology, Davis, CA 95616 USA
 FEATURES
 source
 1. 234
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="4"
 /sex="female"
 /cell_line="HeLa"
 /cell_type="epithelial cell"

BASE COUNT precursor RNA 1. 234
 56 a 60 c 83 g 35 t

ORIGIN

Query Match 100.0%; Score 18; DB 93; Length 234;
 Best Local Similarity 100.0%; Pred. No. 0.73;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcggagcttcagctgag 18
 |||
 Db 211 GCGCGAGCTTCAGCTGAG 228

RESULT 6
 HSA248384 235 bp DNA STS 06-JAN-2000
 LOCUS Homo sapiens STS NRL-402R, sequence tagged site.
 DEFINITION AJ248384
 ACCESSION AJ248384.1 GI:5514722
 VERSION
 KEYWORDS STS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 235)
 AUTHORS Sulimova,G.E., Udina,I.G., Kunizheva,S.S. and Kompanil'tzev,A.A.
 TITLE Creating NotI-STS Markers for Human Chromosome 3
 JOURNAL Mol. Biol. 33, 698-703 (1999)
 REFERENCE 2 (bases 1 to 235)
 AUTHORS Sulimova,G.E.
 TITLE Direct Submission
 JOURNAL Submitted (23-JUN-1999) Sulimova G.E., Laboratory of Comparative
 Animal Genetics, N.I. Vavilov Institute of General Genetics RAS,
 Gubkin Str. 3, Moscow B-333, 117809 GSP-1, RUSSIA
 COMMENT The STS markers registered were developed
 to clones from NotI library of human chromosome
 3 received by E.R. Zbarovskiy (Karolinska
 Institute, Sweden) as a result of collaborative
 research work with Engelhardt Molecular Biology
 Institute of the Russian Academy of Sciences.
 The NotI clones were used to construct NotI map
 of human chromosome 3 by FISH and mapping by a
 somatic cell hybrid panel.
 FEATURES
 source
 1. 235
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="3"
 /cell_type="mouse/human microcell hybrid lines"
 /cell_line="WCH903.1"
 /map="3p21.3"
 /clone_lib="NotI-linking library"
 1. 21
 /note="NRL-402R forward primer"
 /PCR_conditions="95deg. 0.6min, 57.5deg. 0.6min, 72deg.
 1.0min
 22. 215
 /standard_name="NRL-402R"
 /note="STS PCR product"

```
BASE COUNT      60 a      69 c      78 g      37 t
ORIGIN          /rpt_family="tarsier Alu"
                /rpt_type=dispersed
```

Query Match 100.0%; Score 18; DB 88; Length 244;
 Best Local Similarity 100.0%; Pred. No. 0.72;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggcggagcttcagtgag 18
 |||
 DB 191 GCGGAGCTTCAGCTGAG 208

RESULT 10.
 HSI1MDP1 245 bp DNA PRI 22-APR-1997
 LOCUS H.sapiens P1 promoter sequence of gene encoding inosine
 DEFINITION monophosphate dehydrogenase I.
 ACCESSION Y08944.1 GI:1628396
 VERSION Y08944.1 GI:1628396
 KEYWORDS inosine monophosphate dehydrogenase.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 245)
 AUTHORS Gu,J.J., Spychala,J. and Mitchell,B.S.
 TITLE Regulation of the human inosine monophosphate dehydrogenase type I
 JOURNAL J. Biol. Chem. 272 (7), 4458-4466 (1997)
 MEDLINE 97172526
 REFERENCE 2 (bases 1 to 245)
 AUTHORS Gu,J.
 TITLE Direct Submission
 JOURNAL Submitted (18-OCT-1996) J. Gu, University of North Carolina-Chapel
 Hill, Pharmacology, 1106 P10b CB7365, North Carolina 27599, USA
 FEATURES
 source
 1..245
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="7"
 /cell_type="leukocyte"
 /map="q31.3-q32"
 /clone="phage 1111a"
 /dev_stage="adult"
 1..245
 /gene="inosine monophosphate dehydrogenase type I"
 1..245
 /gene="inosine monophosphate dehydrogenase type I"
 /note="P1"
 BASE COUNT 60 a 65 c 81 g 39 t
 ORIGIN

Query Match 100.0%; Score 18; DB 93; Length 245;
 Best Local Similarity 100.0%; Pred. No. 0.72;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggcggagcttcagtgag 18
 |||
 DB 188 GCGGAGCTTCAGCTGAG 205

RESULT 11.
 AF270578 248 bp DNA PRI 09-AUG-2000
 LOCUS Homo sapiens clone 18ptelc6.59t3 sequence.
 DEFINITION AF270578
 ACCESSION AF270578
 VERSION AF270578.1 GI:9755120
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 248)

AUTHORS Riethman,H.C. and Moyzis,R.K.
 TITLE Integration of telomeric DNA sequences with the human reference
 sequence
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 248)
 AUTHORS Riethman,H.C. and Moyzis,R.K.
 TITLE Direct Submission
 JOURNAL Submitted (19-MAY-2000) Molecular Genetics, The Wistar Institute,
 3601 Spruce St., Philadelphia, PA 19104, USA
 FEATURES
 source
 1..248
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="18"
 /map="18ptel"
 /clone="18ptelc6.59t3"
 BASE COUNT 35 a 74 c 55 g 84 t
 ORIGIN

Query Match 100.0%; Score 18; DB 89; Length 248;
 Best Local Similarity 100.0%; Pred. No. 0.72;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggcggagcttcagtgag 18
 |||
 DB 155 GCGGAGCTTCAGCTGAG 138

RESULT 12
 HUMALCE21 249 bp ss-RNA PRI 15-APR-1994
 LOCUS Human carcinoma cell-derived Alu RNA transcript, clone CE21.
 DEFINITION M87925
 ACCESSION M87925.1 GI:174872
 VERSION M87925.1 GI:174872
 KEYWORDS Alu repeat.
 SOURCE Homo sapiens male embryo carcinoma cDNA to other RNA.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 249)
 AUTHORS Sinnett,D., Richey,C., Deragon,J.-M. and Labuda,D.
 TITLE Alu RNA transcripts in human embryonal carcinoma cells. Model of
 post-transcriptional selection of master sequences
 JOURNAL J. Mol. Biol. (1992) In press
 FEATURES
 source
 1..249
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /cell_line="NTERA2D1"
 /dev_stage="embryo"
 /sex="male"
 /tissue_type="carcinoma"
 BASE COUNT 66 a 66 c 80 g 37 t
 ORIGIN

Query Match 100.0%; Score 18; DB 97; Length 249;
 Best Local Similarity 100.0%; Pred. No. 0.72;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggcggagcttcagtgag 18
 |||
 DB 173 GCGGAGCTTCAGCTGAG 190

RESULT 13
 HS31B4F 257 bp DNA PRI 19-OCT-1995
 LOCUS H.sapiens CpG island DNA genomic MseI fragment, clone 31b4, forward
 DEFINITION read cpg31b4.ftla.
 ACCESSION Z60686
 VERSION Z60686.1 GI:1032790

KEYWORDS CPG island; genomic MseI fragment.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS 1 (bases 1 to 257)

TITLE Macdonald, M., Huckle, E., Wilkinson, P. and Mickle, G.

JOURNAL Direct Submission

Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire, CB10 1RQ, England. E-mail contact: humquerry@sanger.ac.uk

2 (bases 1 to 257)

REFERENCE Cross, S.H., Charlton, J.A., Nan, X. and Bird, A.P.

AUTHORS Purification of CPG islands using a methylated DNA binding column

TITLE Nat. Genet. 6 (3), 236-244 (1994)

JOURNAL 94282070

MEDLINE

COMMENT Vector: pGEM-52f(-)

Clones are available from the UK MRC Human Genome Mapping Project

Resourse Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:

http://www.hgmp.mrc.ac.uk/ for details

or contact: biohelpehgmp.mrc.ac.uk.

FEATURES

source

1..257

Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

/sex="male"

/dev_stage="adult"

/tissue_type="blood"

/clone_lib="CGI-1"

/clone="31d4"

BASE COUNT 43 a 87 c 70 g 55 t 2 others

ORIGIN

Query Match 100.0%; Score 18; DB 92; Length 257;

Best Local Similarity 100.0%; Pred. No. 0.72;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggcggagcttcagtcag 18

|||||

Db 72 GCGGAGCTTCAGTCAG 55

|||||

RESULT 14

G38317 262 bp DNA STS 02-JUN-1998

LOCUS RPI1-4-790D12T7 Human Homo sapiens STS genomic, sequence tagged

DEFINITION site.

ACCESSION G38317

VERSION G38317.1 GI:3172222

KEYWORDS STS.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 262)

AUTHORS de Jong, P.J.

TITLE Human chromosome 14q24.3

JOURNAL Unpublished (1998)

COMMENT

Contact: Pieter J. de Jong

Human Genetics Department

Roswell Park Cancer Institute

Elm and Carlton Streets, Buffalo, NY, 14263, USA

Tel: 716-845-3168

Fax: 716-845-8449

Email: pieter@dejong.med.buffalo.edu

Primer A: TATTTCTTAAAGGAGAGGCTTC

Primer B: TCAAGCTCTGATTCATTAAT

STS size: 81

PCR Profile:

Presoak: 0 degrees C for 0.00 minute(s)

Denaturation 94 degrees C for 30 second(s)

Annealing 53.7 degrees C for 30 second(s)

Polymerization: 72 degrees C for 1.00 minute(s)

PCR Cycles: 35

Thermal Cycler: Perkin Elmer TC

Protocol:

Template: 30-100 ng

Primer: each 1µM

dNTPs: each 200µM

Tag Polymerase: 0.05 units/µl

Total Vol: 10 µl

Buffer:

MgCl2: 1.5mM

KCl: 50 mM

Tris-HCl 10 mM

pH: 8.3.

FEATURES

source

1..262

Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="Human"

STS

primer_bind 68..148

primer_bind 68..89

BASE COUNT 82 a 60 c 66 g 54 t

ORIGIN

Query Match 100.0%; Score 18; DB 54; Length 262;

Best Local Similarity 100.0%; Pred. No. 0.72;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggcggagcttcagtcag 18

|||||

Db 175 GCGGAGCTTCAGTCAG 192

|||||

RESULT 15

HSU14706 262 bp DNA PRI 11-MAR-1995

LOCUS Human Alu-Sb2 repeat, clone HUM-10.

DEFINITION U14706

ACCESSION U14706.1 GI:551115

VERSION U14706.1 GI:551115

KEYWORDS SINE.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (sites)

AUTHORS Zietkiewicz, E., Richer, C., Makalowski, W., Jurta, J. and Labuda, D.

TITLE A young Alu subfamily amplified independently in human and African

JOURNAL great apes lineages

MEDLINE Nucleic Acids Res. 22 (25), 5608-5612 (1994)

REFERENCE 95140622

2 (bases 1 to 262)

AUTHORS Makalowski, W.

TITLE Direct Submission

JOURNAL Submitted (15-SEP-1994) Wojciech Makalowski, National Center for

Biotechnology Information, NLM, NIH, 8600 Rockville Pike, Bethesda,

MD 20894, USA

FEATURES

source

1..262

Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="HUM-10"

1..16

primer_bind 1..262

repeat_region

/rpt_type=tandem

/evidence=experimental

/rpt_family="Alu-Sb2"

/rpt_unit=1..262

/complement(244..262)

BASE COUNT 62 a 69 c 89 g 42 t

ORIGIN

Thu Nov 15 11:23:47 2001

us-09-663-020-8.rge

Page 7

Query Match 100.0%; Score 18; DB 97; Length 262;
Best Local Similarity 100.0%; Pred. No. 0.72;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ggcgagcttcagtgag 18
|||||
Db 211 ggcgagcttcagtgag 228

Search completed: November 15, 2001, 06:17:28
Job time: 8013 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 15, 2001, 04:03:55 ; Search time 1266.25 Seconds
(without alignments)
293.170 Million cell updates/sec

Title: US-09-663-020-7
Perfect score: 24
Sequence: 1 tcagattatttgggttcacac 24

Scoring table: OLIGO-MUC
Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 7733874588 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:
1: gb_ba1:
2: gb_ba2:
3: gb_ba3:
4: gb_in1:
5: gb_in2:
6: gb_in3:
7: gb_om:
8: gb_ov:
9: gb_pat1:
10: gb_pat2:
11: gb_ph:
12: gb_pl1:
13: gb_pl2:
14: gb_pl3:
15: gb_pl4:
16: em_ba1:
17: em_ba2:
18: em_fun:
19: em_htgo_hum:
20: em_htgo_inv:
21: em_htgo_rod:
22: em_htg_hum1:
23: em_htg_hum2:
24: em_htg_hum3:
25: em_htg_hum4:
26: em_htg_hum5:
27: em_htg_hum6:
28: em_htg_hum7:
29: em_htg_hum8:
30: em_htg_inv1:
31: em_htg_inv2:
32: em_htg_other:
33: em_htg_rod:
34: em_hum1:
35: em_hum2:
36: em_hum3:
37: em_hum4:
38: em_hum5:
39: em_hum6:
40: em_hum7:
41: em_in:
42: em_om:
43: em_or:

44: em_ov:
45: em_pat:
46: em_ph:
47: em_pl:
48: em_rod:
49: em_sts:
50: em_sy:
51: em_un:
52: em_v1:
53: gb_sts1:
54: gb_sts2:
55: gb_sts3:
56: gb_sy:
57: gb_un:
58: gb_v11:
59: gb_v12:
60: gb_htg1:
61: gb_htg2:
62: gb_htg3:
63: gb_htg4:
64: gb_htg5:
65: gb_htg6:
66: gb_htg7:
67: gb_htg8:
68: gb_htg9:
69: gb_htg10:
70: gb_htg11:
71: gb_htg12:
72: gb_htg13:
73: gb_htg14:
74: gb_htg15:
75: gb_htg16:
76: gb_htg17:
77: gb_htg18:
78: gb_htg19:
79: gb_htg20:
80: gb_htg21:
81: gb_htg22:
82: gb_htg23:
83: gb_htg24:
84: gb_htg25:
85: gb_pr1:
86: gb_pr2:
87: gb_pr3:
88: gb_pr4:
89: gb_pr5:
90: gb_pr6:
91: gb_pr7:
92: gb_pr8:
93: gb_pr9:
94: gb_r01:
95: gb_r02:
96: gb_in4:
97: gb_pr10:
98: em_ba3:

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	24	100.0	196848	86	AC007684
2	18	75.0	49306	88	AC022237
3	18	75.0	63031	77	AC087465
4	18	75.0	161492	64	AC016693
5	18	75.0	170264	88	AC024589
6	18	75.0	202915	72	AC060773
7	17	70.8	141292	78	AF235106
8	17	70.8	165633	86	AC006924
					Homo sapi

C	9	17	70.8	239254	76	AC079569	AC079569 Mus musculus
C	10	17	70.8	253746	76	AC079519	AC079519 Mus musculus
C	11	16	66.7	1233	8	TRCPMH	DBS857 Theragra ch
C	12	16	66.7	3923	8	AB017819	AB017819 Theragra
C	13	16	66.7	6298	94	AF302668	AF302668 Mus musculus
C	14	16	66.7	11199	1	AEO01138	AEO01138 Borrelia
C	15	16	66.7	23091	65	AC017994	AC017994 Drosophila
C	16	16	66.7	23611	63	AC015065	AC015065 Drosophila
C	17	16	66.7	38201	85	AC000004	AC000004 Homo sapi
C	18	16	66.7	38861	85	AC000038	AC000038 Homo sapi
C	19	16	66.7	46335	86	AC005744	AC005744 Homo sapi
C	20	16	66.7	79001	63	AC015386	AC015386 Drosophila
C	21	16	66.7	110302	83	AP002013	AP002013 Homo sapi
C	22	16	66.7	121474	60	AC007150	AC007150 Drosophila
C	23	16	66.7	125057	89	AL133419	AL133419 Human DNA
C	24	16	66.7	132010	68	AC024351	AC024351 Homo sapi
C	25	16	66.7	146244	83	AP002083	AP002083 Homo sapi
C	26	16	66.7	146438	93	HSBA1911L9	AL117329 Human DNA
C	27	16	66.7	154090	88	AC025577	AC025577 Homo sapi
C	28	16	66.7	154158	80	AL365237	AL365237 Homo sapi
C	29	16	66.7	155569	70	AC026862	AC026862 Homo sapi
C	30	16	66.7	156150	69	AC026002	AC026002 Homo sapi
C	31	16	66.7	158340	88	AC073109	AC073109 Homo sapi
C	32	16	66.7	159173	64	AC015992	AC015992 Homo sapi
C	33	16	66.7	159424	79	AL161614	AL161614 Homo sapi
C	34	16	66.7	162832	4	AC013584	AC013584 Homo sapi
C	35	16	66.7	163162	4	AC006247	AC006247 Drosophila
C	36	16	66.7	163391	73	AC067805	AC067805 Homo sapi
C	37	16	66.7	165565	88	AC073651	AC073651 Homo sapi
C	38	16	66.7	167388	73	AC068969	AC068969 Homo sapi
C	39	16	66.7	167878	72	AC055821	AC055821 Homo sapi
C	40	16	66.7	168031	73	AC068430	AC068430 Homo sapi
C	41	16	66.7	168111	69	AC025982	AC025982 Homo sapi
C	42	16	66.7	168134	88	AC021058	AC021058 Homo sapi
C	43	16	66.7	168880	66	AC021654	AC021654 Homo sapi
C	44	16	66.7	170962	80	AL389927	AL389927 Homo sapi
C	45	16	66.7	172326	68	AC023965	AC023965 Homo sapi

ALIGNMENTS

AC079569	Mus musculus
AC079519	Mus musculus
D85857	Therapsia ch
AB017819	Therapsia
AF302688	Mus musculus
AE001138	Borrelia
AC017994	Drosophila
AC015065	Drosophila
AC000046	Mus musculus
AC000003	Mus musculus
AC005744	Mus musculus
AC015386	Drosophila
AP002013	Mus musculus
AC007150	Drosophila
AC033419	Human DNA
AC024351	Mus musculus
AP002083	Mus musculus
AC017329	Human DNA
AC025577	Mus musculus
AL365257	Mus musculus
AC026662	Mus musculus
AC026602	Mus musculus
AC073109	Mus musculus
AC015992	Mus musculus
AL161614	Mus musculus
AC013584	Mus musculus
AC006247	Drosophila
AC067365	Mus musculus
AC073801	Mus musculus
AC068969	Mus musculus
AC058821	Mus musculus
AC068430	Mus musculus
AC025082	Mus musculus
AC021058	Mus musculus
AC021654	Mus musculus
AL389927	Mus musculus
AC023965	Mus musculus

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

MO 63108, USA
5 (bases 1 to 196848)
Waterston, R.
Direct Submission
Submitted (21-DEC-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Sep 8, 1999 this sequence version replaced g1:5001493.

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics -----
Center project name: H_NH0449616

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >30): an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION: Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION: The RPC1-11 human BAC library was made from the blood of one male donor, as described by Osoegwa, K., Moon, P. Y., Zhao, B., Frangen, E., Tatenno, M., Cattnese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Fleiter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is Rp11-395B14. Actual start of this clone is at base position 1 of Rp11-449G16, actual end is at base position 196848 of Rp11-449G16.

FEATURES
source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-449G16"
/clone.lib="RPIC11"
1..153
/rpc_family="L1"
154..285
/rpc_family="Alu"
286..319
/rpc_family="L1"
320..356
/rpc_family="(TG)n"
357..423
/rpc_family="L1"
456..746
/rpc_family="L1"
747..1036
/rpc_family="Alu"
1037..1882
/rpc_family="L1"

```

REFERENCE 4 (bases 1 to 196648)
AUTHORS Waterston, R. H.
TITLE Direct Submission
JOURNAL Submitted (08-SEP-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, Missouri 63110

```

repeat_region 1962..2006 /rpt_family="AT_rich"
repeat_region 2007..2119 /rpt_family="L1"
repeat_region 2120..2422 /rpt_family="Alu"
repeat_region 2423..2568 /rpt_family="Alu"
misc_feature 3822..4005 /rpt_family="L1"
repeat_region /note="match to EST A1635735 (NID:g4687065) t286g12.x1"
4070..4364 /rpt_family="Alu"
misc_feature 4099..4528 /note="match to EST AA553666 (NID:g2324205) nk79g10.s1"
4631..4737 /rpt_family="L2"
repeat_region 4872..5166 /rpt_family="L2"
misc_feature 4894..4992 /rpt_family="Alu"
repeat_region /note="match to EST AA553666 (NID:g2324205) nk79g10.s1"
5167..5187 /rpt_family="(TAA)n"
5188..5198 /rpt_family="Alu"
repeat_region 5623..6021 /rpt_family="Alu"
repeat_region 6057..6200 /rpt_family="L1"
repeat_region 6403..6484 /rpt_family="Alu"
repeat_region 6523..6631 /rpt_family="L2"
repeat_region 6839..6853 /rpt_family="MIR"
repeat_region 6834..6876 /rpt_family="Alu"
repeat_region /rpt_family="(ATTC)n"
6877..7160 /rpt_family="Alu"
repeat_region 8456..8750 /rpt_family="Alu"
repeat_region 8843..9134 /rpt_family="Alu"
repeat_region 9311..9340 /rpt_family="Alu"
repeat_region /rpt_family="(TTTC)n"
9341..9609 /rpt_family="Alu"
repeat_region 9619..9924 /rpt_family="Alu"
repeat_region 9932..10234 /rpt_family="Alu"
repeat_region 10766..11251 /rpt_family="Alu"
repeat_region /rpt_family="MER2_type"
11252..11544 /rpt_family="Alu"
repeat_region 11545..11649 /rpt_family="MER2_type"
11650..11737 /rpt_family="Alu"
repeat_region 11738..11941 /rpt_family="Alu"
repeat_region 11942..12061 /rpt_family="MER2_type"
repeat_region 12062..12508 /rpt_family="Alu"
repeat_region /rpt_family="MER2_type"
12927..12973 /rpt_family="MERL_type?"
repeat_region 13277..13396 /rpt_family="L2"
repeat_region 13424..13788 /rpt_family="Retroviral"
repeat_region 14057..14358 /rpt_family="L2"
repeat_region 14399..14503

```

```

/rpt_family="L2"
repeat_region 14781..14906 /rpt_family="MERL_type"
repeat_region 15100..15392 /rpt_family="Alu"
repeat_region 15605..15974 /rpt_family="Alu"
repeat_region 15999..16380 /rpt_family="MERL"
repeat_region 16454..16763 /rpt_family="L2"
repeat_region 16784..16827 /rpt_family="Alu"
repeat_region 16873..16951 /rpt_family="MIR"
repeat_region 17102..17213 /rpt_family="MIR"
repeat_region 17237..17531 /rpt_family="L1"
repeat_region 17532..17559 /rpt_family="Alu"
repeat_region 18102..18408 /rpt_family="AT_rich"
repeat_region 18481..18576 /rpt_family="Alu"
repeat_region /rpt_family="L2"

```

Query Match 100.0%; Score 24; DB 86; Length 196848;
 Best Local Similarity 100.0%; Pred. No. 0.0034;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcagattatttggcctcacc 24
 Db 112853 TCAGATTATTATTGGCCTCACGC 112876

```

RESULT 2
AC022237 49306 bp DNA PRI 15-DEC-2000
LOCUS Homo sapiens chromosome 5 clone RP11-323c17 T7 end, complete
DEFINITION
ACCESSION AC022237
VERSION AC022237.17 GI:11863032
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 49306)
Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
Fedorispiet,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
Mao,J., Lam,B., Marathe,R., Miranda,M., Morehouse,A.J., Nguyen,M.,
Oefner,P., Palm,C.J., Ramirez,D., Southwick,A.M., Wilhelm,J.,
Yu,S. and Davis,R.W.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 49306)
Bruno,D., Conn,L., Dela Rosa,M., Fedorispiet,N., Foreman,P.,
Glukhov,S., Hansen,N., Hyman,R., Mao,J., Marathe,R.,
Morehouse,A.J., Oefner,P., Palm,C.J., Ramirez,D., Wilhelm,J.,
Yu,S. and Davis,R.W.
JOURNAL Direct Submission
TITLE Submitted (27-JAN-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
REFERENCE 3 (bases 1 to 49306)
Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
Fedorispiet,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
Lam,B., Mao,J., Marathe,R., Miranda,M., Morehouse,A.J., Nguyen,M.,
Oefner,P., Palm,C.J., Ramirez,D., Southwick,A.M., Wilhelm,J.,
Yu,S. and Davis,R.W.
JOURNAL Direct Submission
TITLE Submitted (15-DEC-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA

```

COMMENT On Dec 15, 2000 this sequence version replaced gi:11095463.
 ----- Genome Center
 Center: Stanford DNA Sequencing and Technology Development
 Center
 Center code: SDBTDC
 Web site: <http://sequence-www.stanford.edu/group/human/>
 Contact: hum-info@sequence.stanford.edu
 ----- Project Information
 Center project name: 733
 Center clone name: RP11-323C17

----- Summary
 This sequence was finished as follows unless otherwise noted:
 all regions were double stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30);
 an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by sequence
 from more than one subclone; and the assembly was confirmed by
 restriction digest.
 This sequence is only the first 49306 bases (77 end) of 323C17.
 The finished clone CTD-2203L19 (AC025771) overlaps this sequence
 beginning at base 10815.
 Location/Qualifiers

FEATURES
 Source
 1. 49306
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="5"
 /clone="RP11-323C17"
 /clone_11b="RPC1 human BAC library 11"
 BASE COUNT 16374 a 8614 c 8562 g 15756 t
 ORIGIN

Query Match 75.0%; Score 18; DB 88; Length 49306;
 Best Local Similarity 100.0%; Pred. No. 6.5;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 2 cagattatttggcctt 19
 |||||
 DB 29550 CAGATTATTGGCCTT 29567

RESULT 3
 AC087465
 LOCUS AC087465 63031 bp DNA HTG 12-APR-2001
 DEFINITION Homo sapiens chromosome 15 clone RP11-94P14 map 15, LOW-PASS
 SEQUENCE SAMPLING.
 AC087465
 VERSION AC087465.2 GI:13605992
 KEYWORDS HTG: HTGS_PHASEO.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 63031)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Homo sapiens chromosome 15, clone RP11-94P14
 Unpublished
 2 (bases 1 to 63031)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
 Bana,N., Bastien,V., Boguslavsky,L., Boukhalter,B., Brown,A.,
 Camarata,A., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
 Collymore,A., Cooke,P., Dearrellano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Fato,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,
 Gaidyna,S., Ginde,S., Coyette,M., Graham,C., Grand-Pierre,N.,
 Hagos,B., Heaford,A., Horton,L., Hulme,M., Iliev,I., Johnson,R.,
 Jones,C., Karatas,A., Laroque,K., Lamazares,R., Landers,T.,
 Lehocsky,J., Levine,R., Liu,G., Maclean,C., Macdonald,P.,
 Marquis,N., Mathews,C., McCarthy,M., McKean,P., McKernan,K.,
 McPheters,R., Melgrim,J., Meneus,L., Mihova,T., Mlenga,V.,
 Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
 Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,

TITLE JOURNAL COMMENT

Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
 Roy,A., Santos,R., Schauer,S., Schupack,R., Seaman,S., Severy,P.,
 Sounguez,C., Spencer,B., Stange-Thoman,N., Stojanovic,N.,
 Strauss,C., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
 Travers,M., Travis,N., Triggillo,J., Vassiliev,H., Viet,R., Vo,A.,
 Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
 Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (03-JAN-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Apr 12, 2001 this sequence version replaced gi:12039478.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www.seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: 111896
 Center clone name: 94_P_14

* NOTE: This record contains 76 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1 719: contig of 719 bp in length
 * 820 819: gap of 100 bp
 * 820 1577: contig of 758 bp in length
 * 1578 1677: gap of 100 bp
 * 1678 2431: contig of 754 bp in length
 * 2432 2531: gap of 100 bp
 * 2532 3282: contig of 751 bp in length
 * 3283 3382: gap of 100 bp
 * 3383 4099: contig of 717 bp in length
 * 4100 4199: gap of 100 bp
 * 4200 4856: contig of 657 bp in length
 * 4857 4956: gap of 100 bp
 * 4957 5668: contig of 712 bp in length
 * 5669 5768: gap of 100 bp
 * 5769 6478: contig of 710 bp in length
 * 6479 6578: gap of 100 bp
 * 6579 7284: contig of 706 bp in length
 * 7285 7384: gap of 100 bp
 * 7385 8133: contig of 749 bp in length
 * 8134 8233: gap of 100 bp
 * 8234 8982: contig of 749 bp in length
 * 8983 9082: gap of 100 bp
 * 9083 9832: contig of 750 bp in length
 * 9833 9932: gap of 100 bp
 * 9933 10659: contig of 727 bp in length
 * 10660 10759: gap of 100 bp
 * 10760 11477: contig of 718 bp in length
 * 11478 11577: gap of 100 bp
 * 11578 12297: contig of 720 bp in length
 * 12298 12397: gap of 100 bp
 * 12398 13126: contig of 729 bp in length
 * 13127 13226: gap of 100 bp
 * 13227 13962: contig of 736 bp in length
 * 13963 14062: gap of 100 bp
 * 14063 14802: contig of 740 bp in length
 * 14803 14902: gap of 100 bp
 * 14903 15649: contig of 747 bp in length
 * 15650 15749: gap of 100 bp
 * 15750 16456: contig of 707 bp in length

```

* 16457 16556: gap of 100 bp
* 16557 17307: contig of 751 bp in length
* 17308 17407: gap of 100 bp
* 17408 18156: contig of 749 bp in length
* 18157 18256: gap of 100 bp
* 18257 19009: contig of 753 bp in length
* 19010 19109: gap of 100 bp
* 19110 19815: contig of 706 bp in length
* 19816 19915: gap of 100 bp
* 19916 20628: contig of 713 bp in length
* 20629 20728: gap of 100 bp
* 20729 21435: contig of 707 bp in length
* 21436 21535: gap of 100 bp
* 21536 22260: contig of 725 bp in length
* 22261 22360: gap of 100 bp
* 22361 23061: contig of 701 bp in length
* 23062 23161: gap of 100 bp
* 23162 23898: contig of 737 bp in length
* 23899 23998: gap of 100 bp
* 23999 24746: contig of 748 bp in length
* 24747 24846: gap of 100 bp
* 24847 25596: contig of 750 bp in length
* 25597 25696: gap of 100 bp
* 25697 26417: contig of 721 bp in length
* 26418 26517: gap of 100 bp
* 26518 27234: contig of 717 bp in length
* 27235 27334: gap of 100 bp
* 27335 28051: contig of 717 bp in length
* 28052 28151: gap of 100 bp
* 28152 28884: contig of 733 bp in length
* 28885 28984: gap of 100 bp
* 28985 29718: contig of 734 bp in length
* 29719 29818: gap of 100 bp
* 29819 30551: contig of 733 bp in length
* 30552 30651: gap of 100 bp
* 30652 31376: contig of 725 bp in length
* 31377 31476: gap of 100 bp
* 31477 32217: contig of 741 bp in length
* 32218 32317: gap of 100 bp
* 32318 33063: contig of 746 bp in length
* 33064 33163: gap of 100 bp
* 33164 33876: contig of 713 bp in length
* 33877 33976: gap of 100 bp
* 33977 34724: contig of 748 bp in length
* 34725 34824: gap of 100 bp
* 34825 35560: contig of 736 bp in length
* 35561 35660: gap of 100 bp
* 35661 36415: contig of 755 bp in length
* 36416 36515: gap of 100 bp
* 36516 37244: contig of 729 bp in length
* 37245 37344: gap of 100 bp
* 37345 38043: contig of 699 bp in length
* 38044 38143: gap of 100 bp
* 38144 38869: contig of 726 bp in length
* 38870 38969: gap of 100 bp
* 38970 39711: contig of 742 bp in length
* 39712 39811: gap of 100 bp
* 39812 40543: contig of 732 bp in length
* 40544 40643: gap of 100 bp
* 40644 41386: contig of 743 bp in length
* 41387 41486: gap of 100 bp
* 41487 42228: contig of 742 bp in length
* 42229 42328: gap of 100 bp
* 42329 42908: contig of 580 bp in length
* 42909 43008: gap of 100 bp
* 43009 43738: contig of 730 bp in length
* 43739 43838: gap of 100 bp
* 43839 44578: contig of 740 bp in length
* 44579 44678: gap of 100 bp
* 44679 45432: contig of 754 bp in length
* 45433 45532: gap of 100 bp
* 45533 46269: contig of 737 bp in length
* 46270 46369: gap of 100 bp

```

```

* 46370 47103: contig of 734 bp in length
* 47104 47203: gap of 100 bp
* 47204 47915: contig of 712 bp in length
* 47916 48015: gap of 100 bp
* 48016 48776: contig of 761 bp in length
* 48777 48876: gap of 100 bp
* 48877 49627: contig of 751 bp in length
* 49628 49727: gap of 100 bp
* 49728 50475: contig of 748 bp in length
* 50476 50575: gap of 100 bp
* 50576 51297: contig of 721 bp in length
* 51297 51396: gap of 100 bp
* 51397 52108: contig of 712 bp in length
* 52109 52208: gap of 100 bp
* 52209 52949: contig of 741 bp in length
* 52950 53049: gap of 100 bp
* 53050 53777: contig of 728 bp in length
* 53778 53877: gap of 100 bp
* 53878 54631: contig of 754 bp in length
* 54632 54731: gap of 100 bp
* 54732 55469: contig of 738 bp in length
* 55470 55569: gap of 100 bp
* 55570 56320: contig of 751 bp in length
* 56321 56420: gap of 100 bp
* 56421 57177: contig of 757 bp in length

```

```

Query Match 75.0%; Score 18; DB 77; Length 63031;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 2 cagattatttggcctt 19
Db 36897 CAGATTATTGCGCTT 36914

```

```

RESULT 4
AC016693
LOCUS Homo sapiens chromosome 15 clone RP11-94P14, WORKING DRAFT
DEFINITION AC016693
ACCESSION AC016693.4 GI:7230876
VERSION AC016693.4
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 161492)
REFERENCE Waterston, R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 161492)
REFERENCE Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (04-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Mar 13, 2000 this sequence version replaced gl:6838893.

```

COMMENT

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0094P14
----- Summary Statistics -----
Sequencing vector: M13; 78%
Sequencing vector: plasmid; 22%
Chemistry: Dye-terminator Big Dye; 22% of reads
Chemistry: Dye-terminator Big Dye; 22% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 150210 bases at least Q40
Consensus quality: 153641 bases at least Q30

```

Consensus quality: 155892 bases at least Q20
 Insert size: 15800; agarose-fp
 Insert size: 15982; sum-of-ctrls
 Quality coverage: 4.03 in Q20 bases; agarose-fp
 Quality coverage: 3.98 in Q20 bases; sum-of-ctrls

NOTE: This is a 'working draft' sequence. It currently consists of 17 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1
 1806: contig of 1806 bp in length
 1907: gap of unknown length
 1907: contig of 1994 bp in length
 3901: gap of unknown length
 4001: contig of 2729 bp in length
 6730: gap of unknown length
 6830: contig of 3331 bp in length
 10161: gap of unknown length
 10261: contig of 2372 bp in length
 12633: gap of unknown length
 12733: contig of 6020 bp in length
 18753: gap of unknown length
 18853: gap of unknown length
 25699: contig of 6846 bp in length
 25799: gap of unknown length
 32439: contig of 6641 bp in length
 32540: gap of unknown length
 40119: contig of 7580 bp in length
 40219: gap of unknown length
 53097: contig of 12878 bp in length
 53197: gap of unknown length
 63547: contig of 10350 bp in length
 63548: gap of unknown length
 63648: gap of unknown length
 75921: contig of 12273 bp in length
 76021: gap of unknown length
 87094: contig of 11074 bp in length
 87195: gap of unknown length
 104083: contig of 16889 bp in length
 104184: gap of unknown length
 120980: contig of 16796 bp in length
 121079: gap of unknown length
 143060: contig of 21981 bp in length
 143161: gap of unknown length
 161492: contig of 18332 bp in length.

FEATURES
 source
 1. 161492 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="15"
 /clone="RP11-94P14"
 misc_feature
 1. 1806
 /note="assembly_name:Contig10"
 misc_feature
 1907. 3900
 /note="assembly_name:Contig11"
 misc_feature
 4001. 6729
 /note="assembly_name:Contig12"
 misc_feature
 6830. 10160
 /note="assembly_name:Contig13"
 misc_feature
 10261. 12632
 /note="assembly_name:Contig14"
 misc_feature
 12733. 18752
 /note="assembly_name:Contig15"
 misc_feature
 18853. 25698
 /note="assembly_name:Contig16"
 misc_feature
 25799. 32439
 /note="assembly_name:Contig17"
 misc_feature
 32540. 40119
 /note="assembly_name:Contig18"
 misc_feature
 40220. 53097
 /note="assembly_name:Contig19"

misc_feature 53198. 63547
 /note="assembly_name:Contig20"
 misc_feature 63648. 75920
 /note="assembly_name:Contig21"
 misc_feature 76021. 87094
 /note="assembly_name:Contig22"
 misc_feature 87195. 104083
 /note="assembly_name:Contig23"
 misc_feature 104184. 120979
 /note="assembly_name:Contig24"
 misc_feature 121080. 143060
 /note="assembly_name:Contig25"
 misc_feature 143161. 161492
 /note="assembly_name:Contig26"

BASE COUNT 49888 a 31184 c 30967 g 47845 t 1608 others
 ORIGIN
 vector_side:right"

Query Match 75.0%; Score 18; DB 64; Length 161492;
 Best Local Similarity 100.0%; Pred. No. 6.1;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 cagattatttggcctt 19
 |||

DB 86674 CAGATTATTGGCCTT 86691

RESULT 5
 AC024589/c
 LOCUS Homo sapiens chromosome 5 clone RP11-42L13, complete sequence.
 DEFINITION AC024589

AC024589.4 GI:13173623
 VERSION
 KEYWORDS

HTG.
 SOURCE
 ORGANISM

human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 170264)
 DOE Joint Genome Institute and Stanford Human Genome Center.

JOURNAL
 Direct Submission

REFERENCE
 2 (bases 1 to 170264)

DOE Joint Genome Institute.

AUTHORS
 Direct Submission

JOURNAL
 Submitted (29-FEB-2000) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

3 (bases 1 to 170264)
 DOE Joint Genome Institute and Stanford Human Genome Center.

Direct Submission
 Submitted (01-MAR-2001) DOE Joint Genome Institute, 2800 Mitchell
 Drive, Walnut Creek, CA 94598, USA

On Mar 1, 2001 this sequence version replaced gi:8576139.

Draft Sequence Produced by DOE Joint Genome Institute

www.jgi.doe.gov
 Finishing Completed at Stanford Human Genome Center

www.shgc.stanford.edu
 Quality: Phrap Quality >=40 99.8% of Sequence;

Estimated Total Number of Errors is 0.2.

FEATURES

source

1. 170264

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="5"

/clone="RP11-42L13"

BASE COUNT 55901 a 29650 c 30820 g 53893 t

ORIGIN

Query Match 75.0%; Score 18; DB 88; Length 170264;
 Best Local Similarity 100.0%; Pred. No. 6.1;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 2 cagattatttggcct 19
 ||||||||||||||||
 Db 75268 CAGATTATTGGCCTT 75251

RESULT 6

AC060773

LOCUS AC060773 202915 bp DNA HTG 10-JAN-2001
 DEFINITION Homo sapiens chromosome 4 clone RP11-741G21 map 4, WORKING DRAFT
 SEQUENCE, 8 unordered pieces.

AC060773

AC060773.3 GI:12061518

HTG: HTGS_PHASE1; HTGS_DRAFT.

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

* consists of 8 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 3959: contig of 3959 bp in length
 * 3960 4059: gap of 100 bp
 * 4060 5548: contig of 1489 bp in length
 * 5549 5648: gap of 100 bp
 * 5649 7747: contig of 2099 bp in length
 * 7748 7847: gap of 100 bp
 * 7848 7565: contig of 67818 bp in length
 * 7566 75765: gap of 100 bp
 * 75766 109526: contig of 33761 bp in length
 * 109527 109626: gap of 100 bp
 * 109627 144431: contig of 34805 bp in length
 * 144432 144531: gap of 100 bp
 * 144532 184185: contig of 39654 bp in length
 * 184186 184285: gap of 100 bp
 * 184286 202915: contig of 18630 bp in length.

FEATURES
 source
 1. 202915
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="4"
 /map="4"
 /clone="RP11-741G21"
 /clone_lib="RPC1-11 Human Male BAC"
 1. 3959
 /note="assembly-fragment
 clone_end:SP6
 vector_side:left"

misc_feature

4060..5548

/note="assembly-fragment"

5649..7747

/note="assembly-fragment"

7848..7565

/note="assembly-fragment"

7566..109526

/note="assembly-fragment"

109627..144431

/note="assembly-fragment"

144532..184185

/note="assembly-fragment"

184286..202915

/note="assembly-fragment"

clone_end:T7

vector_side:right"

BASE COUNT 61574 a 37018 c 37733 g 65883 t 707 others

ORIGIN

Query Match 75.0%; Score 18; DB 72; Length 202915;

Best Local Similarity 100.0%; Pred. No. 6.1;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 cagattatttggcct 19

||||||||||||||||

Db 188527 CAGATTATTGGCCTT 188544

RESULT 7

AF235106

LOCUS AF235106 141292 bp DNA HTG 01-JUN-2000

DEFINITION Homo sapiens chromosome 8 map 8q12.3-8q13.1 clone GSI-110g02.

WORKING DRAFT SEQUENCE, 9 unordered pieces.

AC060773

AF235106.2 GI:8151913

HTG: HTGS_PHASE1; HTGS_DRAFT.

KEYWORDS

SOURCE

ORGANISM

* NOTE: This is a 'working draft' sequence. It currently


```

REFERENCE
AUTHORS      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
              1 (bases 1 to 141292)
TITLE        Schudy,A., Blechschmidt,K., Menzel,U., Polley,A., Reichwald,K.,
              Rump,A., Schilhabel,M.B., Taudien,S., Wen,G., Schlegelberger,B.,
              Siebert,R. and Rosenthal,A.
              Chromosome 8 genomic sequence
              Unpublished
              2 (bases 1 to 141292)
REFERENCE    Schudy,A., Blechschmidt,K., Menzel,U., Polley,A., Reichwald,K.,
              Rosenthal,A., Rump,A., Schilhabel,M.B., Taudien,S. and Wen,G.
              Direct Submission
              Submitted (11-FEB-2000) Genome Analysis, Institute of Molecular
              Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
              3 (bases 1 to 141292)
REFERENCE    Genome Sequencing Center Jena.
              Direct Submission
              Submitted (24-MAY-2000) Genome Analysis, Institute of Molecular
              Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
              On Jun 1, 2000 this sequence version replaced gi:7528258.
COMMENT      * NOTE: This is a 'working draft' sequence. It currently
              * consists of 9 contigs. The true order of the pieces
              * is not known and their order in this sequence record is
              * arbitrary. Gaps between the contigs are represented as
              * runs of N, but the exact sizes of the gaps are unknown.
              * This record will be updated with the finished sequence
              * as soon as it is available and the accession number will
              * be preserved.
              1
              4700: contig of 4700 bp in length
              *
              4801: gap of unknown length
              *
              22393: contig of 17593 bp in length
              *
              22394: gap of unknown length
              *
              22494: 22919: contig of 7426 bp in length
              *
              23920: 30019: gap of unknown length
              *
              30020: 47695: contig of 17676 bp in length
              *
              47696: 47795: gap of unknown length
              *
              47796: 63879: contig of 16084 bp in length
              *
              63880: 63979: gap of unknown length
              *
              63980: 71247: contig of 7268 bp in length
              *
              71248: 71347: gap of unknown length
              *
              71348: 77289: contig of 5942 bp in length
              *
              77290: 77389: gap of unknown length
              *
              77390: 80926: contig of 3537 bp in length
              *
              80927: 81026: gap of unknown length
              *
              81027: 141292: contig of 60266 bp in length.
              *
              Location/Qualifiers
              1. 141292
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /chromosome="8 map Bg12.3-Bg13.1"
              /clone="GSI-110902"
              1. 4700
              /note="assembly-fragment
              clone_end:SP6
              vector_side:left"
              misc_feature 80227..141292
              /note="assembly-fragment
              clone_end:T7
              vector_side:right"
              BASE COUNT 41274 a 27385 c 28529 g 43296 t 808 others
              ORIGIN

Query Match 70.8%; Score 17; DB 78; Length 141292;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 attattttggcctca 21
|||||
DB 31500 ATTATTATGGGCTTCA 31516

RESULT 8

AC006924
LOCUS       AC006924 165633 bp DNA PRI 02-MAR-1999
DEFINITION Homo sapiens, clone hRPK.32_A1, complete sequence.
ACCESSION  AC006924
VERSION    AC006924.3 GI:4314414
KEYWORDS   HTG.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
            1 (bases 1 to 165633)
REFERENCE  Baker,J., Baldwin,J., Barna,N., Beckertly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donegan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heatford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczeky,J., Lien,C., Locke,K., MacDonald,P.,
            Margulis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,D., Mychaleckyj,J.,
            Naylor,J., Nilloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Testafaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
            Direct Submission
            Submitted (25-FEB-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            3 (bases 1 to 165633)
REFERENCE  Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckertly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donegan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heatford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczeky,J., Lien,C., Locke,K., MacDonald,P.,
            Margulis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,D., Mychaleckyj,J.,
            Naylor,J., Nilloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Testafaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
            Direct Submission
            Submitted (02-MAR-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            On Mar 2, 1999 this sequence version replaced gi:4314281.
            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.
            Location/Qualifiers
            1. 165633
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="hRPK.32_A1"
            /clone_id="RPC1-11 human BAC library"
            complement(2..217)
            /rpt_family="MER57-internal"
            complement(220..425)
            /rpt_family="MER57B"
            466..519
            /rpt_family="LTR24B"
            716..1035
            /rpt_family="HERV23"
            1961..2607
            /rpt_family="L1MC1"
            complement(2544..2680)

```

	/note="Single-stranded coverage."		
repeat_region	complement(2608..2873)	repeat_region	22849..22879
repeat_region	/rpt_family="MSTB"	repeat_region	/rpt_family="(TTTG)n"
repeat_region	2874..2898	repeat_region	complement(22880..23161)
repeat_region	/rpt_family="(CGGGGG)n"	repeat_region	/rpt_family="AluY"
repeat_region	complement(2899..3023)	repeat_region	23307..23586
repeat_region	/rpt_family="MSTB"	repeat_region	/rpt_family="LIM4"
repeat_region	3024..3225	repeat_region	complement(25583..26290)
repeat_region	/rpt_family="LIMC1"	repeat_region	/rpt_family="LIMAS"
repeat_region	complement(3981..4126)	repeat_region	26296..26937
repeat_region	/rpt_family="MLTIA2"	repeat_region	/rpt_family="LIMAS"
repeat_region	complement(4237..4340)	repeat_region	26947..27533
repeat_region	/rpt_family="LI"	repeat_region	/rpt_family="LIM4"
repeat_region	4345..4843	repeat_region	complement(27569..27872)
repeat_region	/rpt_family="LIMA1"	repeat_region	/rpt_family="AluY"
repeat_region	complement(4869..4958)	repeat_region	28038..28352
repeat_region	/rpt_family="MLTIA2"	repeat_region	/rpt_family="AluY"
repeat_region	5715..5935	repeat_region	complement(28398..28571)
repeat_region	/rpt_family="HALI"	repeat_region	/rpt_family="LIMB6"
repeat_region	5957..5995	repeat_region	28574..33477
repeat_region	/rpt_family="AT_rich"	repeat_region	/rpt_family="LIM1"
repeat_region	6980..7037	repeat_region	33476..34918
repeat_region	/rpt_family="L2"	repeat_region	/rpt_family="LIMA3"
repeat_region	7183..7269	repeat_region	34936..34975
repeat_region	/rpt_family="AT_rich"	repeat_region	/rpt_family="AT_rich"
repeat_region	complement(7627..7732)	repeat_region	36047..36089
repeat_region	/rpt_family="L2"	repeat_region	/rpt_family="(TTTTA)n"
repeat_region	7871..7914	repeat_region	complement(36099..38002)
repeat_region	/rpt_family="AT_rich"	repeat_region	/rpt_family="LIPAZ"
repeat_region	8166..8347	repeat_region	38445..38484
repeat_region	/rpt_family="LMCS"	repeat_region	/rpt_family="AT_rich"
repeat_region	8386..8693	repeat_region	38502..38758
repeat_region	/rpt_family="LIMCS"	repeat_region	/rpt_family="LTR40a"
repeat_region	9262..9283	repeat_region	complement(38819..39917)
repeat_region	/rpt_family="AT_rich"	repeat_region	/rpt_family="LIMAB"
repeat_region	9412..9452	repeat_region	40334..40402
repeat_region	/rpt_family="AT_rich"	repeat_region	/rpt_family="(CA)n"
repeat_region	9477..9659	repeat_region	40655..40688
repeat_region	/rpt_family="MLTIA2"	repeat_region	/rpt_family="AT_rich"
repeat_region	complement(11356..11518)	repeat_region	41127..41216
repeat_region	/rpt_family="MIR"	repeat_region	/rpt_family="MIR"
repeat_region	14171..14311	repeat_region	41246..41431
repeat_region	/rpt_family="Ricksha"	repeat_region	/rpt_family="LIMB1"
repeat_region	complement(14347..14615)	repeat_region	42317..42422
repeat_region	/rpt_family="AlusX"	repeat_region	/rpt_family="(RTTA)n"
repeat_region	14642..15081	repeat_region	complement(42529..42675)
repeat_region	/rpt_family="Ricksha"	repeat_region	/rpt_family="MER65B"
repeat_region	15082..15258	repeat_region	complement(43006..43358)
repeat_region	/rpt_family="CT_rich"	repeat_region	/rpt_family="MER58B"
repeat_region	15259..15331	repeat_region	43880..43913
repeat_region	/rpt_family="(TTTC)n"	repeat_region	/rpt_family="AT_rich"
repeat_region	15340..15513	repeat_region	44204..44512
repeat_region	/rpt_family="CT_rich"	repeat_region	/rpt_family="AlusX"
repeat_region	complement(15553..15801)	repeat_region	complement(46485..46594)
repeat_region	/rpt_family="AlusX"	repeat_region	/rpt_family="MER94"
repeat_region	15802..16183	repeat_region	complement(46595..46942)
repeat_region	/rpt_family="Ricksha"		
repeat_region	16184..16469		
repeat_region	/rpt_family="Alusq"		
repeat_region	16470..16555		
repeat_region	/rpt_family="(CA)n"		
repeat_region	16556..16574		
repeat_region	/rpt_family="Alusq"		
repeat_region	16575..17130		
repeat_region	/rpt_family="Ricksha"		
repeat_region	complement(17141..17494)		
repeat_region	/rpt_family="LTR33"		
repeat_region	17973..18003		
repeat_region	/rpt_family="(TG)n"		
repeat_region	20928..21613		
repeat_region	/rpt_family="LTR8"		
repeat_region	complement(22686..22815)		
repeat_region	/rpt_family="L2"		

	Query Match	70.8%; Score 17; DB 86; Length 165633;
	Best Local Similarity	100.0%; Pred. No. 21;
	Matches 17; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	5 atttttttgggctcca 21	
DB	41947 ATTATTGTTGGGCTTCA 41963	

RESULT	9	AC079569	239254 bp	DNA	HTG
LOCUS		AC079569/c			
DEFINITION	Mus musculus clone Rp23-71M4, WORKING DRAFT SEQUENCE, 28 unordered				
ACCESSION		AC079569			
VERSION		AC079569.1	GI:9964934		
KEYWORDS		HTG; WGC; BANCAT; WGC; DATA			

SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS 1 (bases 1 to 239254)
JOURNAL DOE Joint Genome Institute.
TITLE Sequencing of Mouse
AUTHORS Unpublished
TITLE 2 (bases 1 to 239254)
AUTHORS DOE Joint Genome Institute.
JOURNAL Direct Submission
TITLE Submitted (02-SEP-2000) Production Sequencing Facility, DOE Joint
AUTHORS Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 1764428
Center clone name: RPCI-23_71M4

Summary Statistics
Consensus quality: 214511 bases at least Q40
Consensus quality: 225489 bases at least Q30
Consensus quality: 227597 bases at least Q20
Estimated insert size: 192000; agarose-fp estimation
Estimated insert size: 236554; sum-of-contigs estimation
Quality coverage: 9.6 in Q20 bases; agarose-fp estimation
Quality coverage: 7.79 in Q20 bases; sum-of-contigs estimation
* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1367: contig of 1367 bp in length
* 1368 1467: gap of unknown length
* 2682: contig of 1215 bp in length
* 2783 2782: gap of unknown length
* 4124 4123: contig of 1341 bp in length
* 4224 4223: gap of unknown length
* 5644 5643: contig of 1420 bp in length
* 5744 7826: gap of unknown length
* 7827 7926: contig of 2083 bp in length
* 7927 10411: gap of unknown length
* 10412 10511: contig of 2485 bp in length
* 10512 12462: gap of unknown length
* 12463 12562: contig of 1951 bp in length
* 12563 14635: gap of unknown length
* 14636 14735: contig of 2073 bp in length
* 14736 18542: gap of unknown length
* 18543 18642: contig of 3807 bp in length
* 18643 23290: gap of unknown length
* 23291 23390: contig of 4648 bp in length
* 23391 27880: gap of unknown length
* 27881 27980: contig of 4490 bp in length
* 27981 32319: gap of unknown length
* 32320 32419: contig of 4339 bp in length
* 32420 35432: gap of unknown length
* 35433 35532: contig of 3013 bp in length
* 35533 41295: gap of unknown length
* 41296 41395: contig of 5763 bp in length
* 41396 48322: gap of unknown length
* 48323 48422: contig of 6927 bp in length
* 48423 54910: gap of unknown length
* 54911 55010: contig of 6488 bp in length
* 55011 63400: gap of unknown length
* 63401 63500: contig of 8390 bp in length
* 63501 74629: gap of unknown length
* 74629: contig of 11129 bp in length

FEATURES
Source
1. 239254
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-71M4"
/clone_1ib="RPCI mouse BAC library 23"
BASE COUNT 68188 a 50089 c 50681 g 67592 t 2704 others
ORIGIN
Query Match 70.8%; Score 17; DB 76; Length 239254;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 3 agattatttggcctt 19
Db 24721 AGATTATTGGCCTT 24705
|||||||
RESULT 10
AC079519
LOCUS AC079519 252746 bp DNA HTG 02-SEP-2000
DEFINITION Mus musculus clone RP23-307N3, WORKING DRAFT SEQUENCE, 32 unordered
pieces
AC079519
VERSION AC079519.1 GI:9964884
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 252746)
JOURNAL DOE Joint Genome Institute.
AUTHORS Sequencing of Mouse
TITLE Unpublished
JOURNAL 2 (bases 1 to 252746)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission.
JOURNAL Submitted (02-SEP-2000) Production Sequencing Facility, DOE Joint
AUTHORS Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 1855075
Center clone name: RPCI-23_307N3

Summary Statistics
Consensus quality: 225531 bases at least Q40
Consensus quality: 238682 bases at least Q30
Consensus quality: 241485 bases at least Q20

Estimated insert size: 206000; agarose-ef estimation
 Estimated insert size: 249646; sum-of-contigs estimation
 Quality coverage: 10.83 in Q20 bases; agarose-ef estimation
 Quality coverage: 8.94 in Q20 bases; sum-of-contigs estimation.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 32 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1      1384: contig of 1384 bp in length
*      1385: gap of unknown length
*      1485: contig of 1503 bp in length
*      2988: gap of unknown length
*      3088: contig of 1090 bp in length
*      4178: gap of unknown length
*      4278: contig of 1068 bp in length
*      5346: gap of unknown length
*      5446: contig of 1142 bp in length
*      6588: gap of unknown length
*      6688: contig of 1323 bp in length
*      8011: gap of unknown length
*      8111: contig of 1168 bp in length
*      9279: gap of unknown length
*      9379: contig of 1023 bp in length
*      10401: gap of unknown length
*      10501: contig of 1146 bp in length
*      10502: gap of unknown length
*      11648: contig of 1167 bp in length
*      11748: gap of unknown length
*      12915: contig of 1167 bp in length
*      13015: gap of unknown length
*      14365: contig of 1351 bp in length
*      14366: gap of unknown length
*      14466: contig of 1996 bp in length
*      16461: gap of unknown length
*      16562: contig of 1129 bp in length
*      17691: gap of unknown length
*      17791: contig of 1448 bp in length
*      19239: gap of unknown length
*      19338: contig of 3984 bp in length
*      23322: gap of unknown length
*      23323: contig of 3153 bp in length
*      23423: gap of unknown length
*      26575: contig of 3143 bp in length
*      26576: gap of unknown length
*      29818: contig of 3174 bp in length
*      29919: gap of unknown length
*      33093: contig of 3174 bp in length
*      33193: gap of unknown length
*      36746: contig of 3553 bp in length
*      36845: gap of unknown length
*      44834: contig of 7989 bp in length
*      44934: gap of unknown length
*      48335: contig of 5742 bp in length
*      49335: gap of unknown length
*      50676: contig of 6083 bp in length
*      50777: gap of unknown length
*      56860: contig of 6016 bp in length
*      62975: gap of unknown length
*      63076: contig of 9530 bp in length
*      72605: gap of unknown length
*      72706: contig of 8730 bp in length
*      81435: gap of unknown length
*      81535: contig of 19008 bp in length
*      100544: gap of unknown length
*      100644: contig of 12946 bp in length
*      113589: gap of unknown length
*      113590: contig of 16305 bp in length
*      129994: gap of unknown length
*      130095: contig of 17211 bp in length
*      147305: gap of unknown length
*      147405: contig of 27379 bp in length
*      174785: gap of unknown length
*      174885: contig of 31934 bp in length

```

```

FEATURES
  source
    * 206819 206818: gap of unknown length
    * 206919 252746: contig of 45828 bp in length.
      Location/Qualifiers
        1..252746
          /organism="Mus musculus"
          /db_xref="taxon:10090"
          /clone="RP23-307N3"
          /clone_11b="Rpci mouse BAC library 23"
BASE COUNT  74071 a 51571 c 52089 g 71910 t 3105 others
ORIGIN

```

```

Query Match      70.8%; Score 17; DB 76; Length 252746;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy      3 agattatttggtcact 19
      |||
Db 223438 ACATTATTGTGGCTT 223454

```

```

RESULT 11
TRCWPBH      1233 bp      mRNA      VRT      06-FEB-1999
LOCUS      Theragra chalcogramma mRNA for myosin head, partial cds.
DEFINITION      D85857
ACCESSION      D85857.1 GI:1396063
KEYWORDS      myosin head.
SOURCE      Theragra chalcogramma skeletal muscle cDNA to mRNA.
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
      Acanthomorphi; Paracanthopterygii; Gadiformes; Gadidae;
      Theragra.

```

```

REFERENCE
  1 (bases 1 to 1233)
  Nishita,K.
  AUTHORS
  TITLE      Submitted (07-JUN-1996) to the DDBJ/EMBL/GenBank databases.
  JOURNAL
  JOURNAL
  TITLE      Kiyoyoshi Nishita, Hokkaido University, Department of Chemistry,
      Faculty of Fisheries: 3-1-1, Minato-cho, Hakodate, Hokkaido
      041-8611, Japan (E-mail:nishitaefish.hokudai.ac.jp,
      Tel:81-138-40-8800, Fax:81-138-40-8800)
  2 (bases 1 to 1233)
  Nishita,K.
  AUTHORS
  TITLE      Primary Structure of Walleye Pollack Myosin Head
  JOURNAL
  JOURNAL
  TITLE      Unpublished (1996)
  FEATURES
    source
      Location/Qualifiers
        1..1233
          /organism="Theragra Chalcogramma"
          /db_xref="taxon:48550"
          /tissue-type="skeletal muscle"
          1..>1233
            /codon_start=1
            /product="myosin head"
            /protein_id="BAA12887.1"
            /db_xref="GI:1396064"
            /translation="MNPXYDKIEDMAMTHINAEVLYNLAERYAAMITYSGLC
      ATVNPYKWLVPYDOSCVNAYRGRKRMHAPPHITSVADNATOFVTDENOSCLITGES
      GAKTVTRKIVFVFAFVTVISGGKADAVAGAKTGSLEQDIIAANPLLEAYGNKTV
      RNNSSRFGKFIIRHFHANGKHSADIEYLLKRSVSPOLPERGYHIFQAMTNK
      PETIEMLTINSNPDPFPCOSQOITVASIDKIELDADDAIDLGFTSEDKVAIFK
      TGAVALHGNKPKOKOREBOAEPDGNEDAKICILYSADMLKALCYPRKVGNY
      VTKGQVYPOVNVNSVALAKSIVYERLPIAMVIRINTMLDTQARQFYIGVLDIAGELF
      DYSMDQLCINFTNERIQO"

```

CDS

```

BASE COUNT  344 a      296 c      301 g      292 t
ORIGIN

```

```

Query Match      66.7%; Score 16; DB 8; Length 1233;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy      8 tatttggtcact 23

```

Db 783 TATTTGGGCTTCACT 798

RESULT 12

AB017819 3923 bp mRNA VRT 30-SEP-1998
 DEFINITION Theragra chalcogramma mRNA for myosin heavy chain, partial cds.
 ACCESSION AB017819
 VERSION AB017819.1 GI:3668186
 KEYWORDS myosin heavy chain.
 SOURCE Theragra chalcogramma skeletal muscle cDNA to mRNA.
 ORGANISM Theragra chalcogramma
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorphi; Paracanthopterygii; Gadiformes; Gadoidi; Gadidae;
 Theragra.

REFERENCE

1 (bases 1 to 3923)
 Nishita, K., Ojima, T. and Watabe, S.
 Direct Submission
 Submitted (22-SEP-1998) to the DDBJ/EMBL/GenBank databases.
 Kiyoshi Nishita, Hokkaido University, Department of Chemistry,
 Faculty of Fisheries, 3-1-1, Minato-cho, Hakodate, Hokkaido
 041-8611, Japan (E-mail: nishita@fish.hokudai.ac.jp,
 Tel: 81-138-40-8800, Fax: 81-138-40-8800)
 2 (sites)
 Ojima, T., Kawashima, N., Inoue, A., Amauchi, A., Togashi, M., Watabe, S.
 and Nishita, K.
 Determination of Primary Structure of Heavy Meromyosin Region of
 Walleye Pollack Myosin Heavy Chain by cDNA Cloning
 Fish. Sci. (1998) in press
 Location/Qualifiers

FEATURES

1..3923
 /organism="Theragra chalcogramma"
 /db_xref="taxon:48550"
 /cfeature="skeletal muscle"
 61..>3921
 /product="myosin heavy chain"
 /protein_id="BA33452.1"
 /db_xref="GI:3668187"
 /translation="MSTDAEMAIYGAAYLRKPERERLEAOSTPEPDAKAAAYADV
 ELVYKCTMRDAGVYVATATKEEDVYPMNPKYDIEMDAMPTILNBS
 VLYNLARYAAMITTYSGLECATVNPYKLPYDSCVNAVYGRKMEAPLIFVS
 DNAPFPLTDRENOSVLTGSGAGVNTKRYIOYFATJANGGCGKADYAGKIKG
 SLEDOIILANPLLEAYGNATRVNDNSRGRKTRIFHANGULSSADISTYLLKRS
 VSPQDPERKTHIFPMATNHPKPEIEMTLITNPDPFPCSGQITVAISIDKEID
 ATDAIDILGFTSEDAVAIFKFGAVLHGNMFKKOREQAPDNGNEADKICYL
 SLNSADMLKALCYPRVAGNEVYKGTQVNVNNSALAKSYERLEFMYRITM
 LDTKQAROFIVGLDIAGFEIPDYSMEOLCINTPEKILQOQFNMFTVEDEYKE
 GIMWETIDFGMDIACILIELEIPMGIESILTEECMPKASDVPFKKLEPQHLKRA
 FEKPKAKKAEAPHSILVHAGVYDVTGMDLNKDPDLDVSTGLYKSSNLLPVL
 YPPVVEVGAGKAGKGGKGGWQVSSQRENLGKLTMTNLRSHPRVRLIPNEKT
 PGLMNLVYHQLKNGVLEGIKICTKGEPRIYADFORVYLVNLSVDPDOFIDN
 KRSEKILSIDVPHDEYKFGHTKFEFFKGLGLTLEMDERKLAALVGMQAGKGY
 MKREYKMTERRAAVYTIQINISFNNVHKPMKYIKIKPLKSAETEKLANKE
 NYDKMTDLAALAKKLEOKKWSILOKNDSTLOMASDNDLNABERCEBLITAK
 IOLEAKVETTERLEDEEELNELLAKKRLDECESEKLDIDLELITAKKELNA
 TEKRYNLTREBASODESVAKLTKKKAQESHOQTLDDIQAEDEKVNLTAKKELA
 QOVDLEGSLEDESKLMDLERKRLLEDLKAETVNDLNDKOOSSEKIKKPF
 TSQHLSEKIDESIGLOKIKLEQARILEELEELAEARAAKAEKRAIDSRILE
 EISERLEAGAGSAQIEMNKREAEFOKRLRLESTOHEATATALKKKODSVAR
 LGEOINLQVAKOLEKSEYKMEIIDLSSNMEVSKAKGNLEKICRALDQLSKIK
 AKSDENAROVNDISARARLITENG"

BASE COUNT 1151 a 854 c 1067 g 851 t
 ORIGIN

Query Match 66.7%; Score 16; DB 8; Length 3923;
 Best Local Similarity 100.0%; Pred. No. 89;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 8 tatttggcttact 23

Db 1083 TATTTGGGCTTCACT 1098

RESULT 13

AF302688 6298 bp DNA ROD 14-DEC-2000
 DEFINITION Mus musculus ventricular myosin regulatory light chain gene,
 promoter, exons 1 and 2, and partial cds.
 ACCESSION AF302688
 VERSION AF302688.1 GI:11761910
 KEYWORDS house mouse.
 SOURCE Mus musculus.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 6298)
 Sanbe, A., Gulick, J. and Robbins, J.
 Ventricular-specific transgenic replacement using a mouse myosin
 light chain promoter
 Unpublished
 2 (bases 1 to 6298)
 Sanbe, A., Gulick, J. and Robbins, J.
 Direct Submission
 Submitted (04-SEP-2000) Molecular Cardiovascular Biology,
 Children's Hospital Research Foundation, 3333 Burnet Avenue,
 Cincinnati, OH 45229, USA
 Location/Qualifiers

FEATURES

1..6298
 /organism="Mus musculus"
 /strain="129/O"
 /db_xref="taxon:10090"
 1..5445
 /number=1
 5446..5489
 join(5446..5489,6209..>6298)
 /product="ventricular myosin regulatory light chain"
 join(5487..5489,6209..>6298)
 /codon_start=1
 /product="ventricular myosin regulatory light chain"
 /protein_id="AAG40240.1"
 /db_xref="GI:11761911"
 /translation="MAPKKAKKIEGGSSVFSNFEQTQIQEKE"
 6209..6298
 /number=2

BASE COUNT 1657 a 1589 c 1520 g 1532 t
 ORIGIN

Query Match 66.7%; Score 16; DB 94; Length 6298;
 Best Local Similarity 100.0%; Pred. No. 87;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 3 agattatttggct 18
 Db 1611 AGATTATTTGGCT 1626

RESULT 14
 AE001138/c
 LOCUS
 DEFINITION Borrelia burgdorferi (section 24 of 70) of the complete genome.
 ACCESSION AE001138 AE000783
 VERSION AE001138.1 GI:2688210
 KEYWORDS Lyme disease spirochete.
 SOURCE Borrelia burgdorferi
 ORGANISM Borrelia burgdorferi
 Bacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia
 burgdorferi group.
 1 (bases 1 to 11199)
 Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R.,
 Lathigra, R., White, O., Ketchum, K.A., Dodson, R., Hickey, E.K.,
 Gwinn, M., Dougherty, B., Tomb, J.F., Fleischmann, R.D., Richardson, D.,

REFERENCE

1 (bases 1 to 11199)
 Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R.,
 Lathigra, R., White, O., Ketchum, K.A., Dodson, R., Hickey, E.K.,
 Gwinn, M., Dougherty, B., Tomb, J.F., Fleischmann, R.D., Richardson, D.,

/product="conserved hypothetical integral membrane protein"
 /protein_id="AAC66694.1"
 /db_xref="GI:2688213"
 /translation="MIFRNSFMALIFSPSTLISYFGDFPQSYIMISMRIPLF
 IMAGIACAKSNLNGNEQIYFGALVYIFSSFLTFNFVFLILSSFFVGLL
 GLIPFTIFPFGKALGLISTGNOLVGFILNMKTGSFNSQTRINSLFALDS
 SLITFLGVSVMLFYVFIHKRTIYGLQLELISNKKITDIFPNIEFFKFFAVGSA
 FLNGLAGSMFVFFRPYLVGLTSGLSGLSSIVAVISGFNVVYVLFSLPSILIEFN
 NFLINIDPKYFVIGLCOSIAIFISLFIKRRKK"
 complement(5734..7194)
 /gene="BB0318"
 complement(5734..7194)
 /note="similar to GB:142023 SP:P44884 PID:1005849
 PID:1220920 PID:1205071 percent identity: 32.53;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="methylgalactoside ABC transporter, ATP-binding
 protein (mg1A)"
 /protein_id="AAC66693.1"
 /db_xref="GI:2688212"
 /translation="MVEFNIVKTFPPIDKPIIDSLINKIGEVKIFTVGKNGEGKST
 LAKIAGLIEFDESELVNGIKQNMVNDKAKNGIYLSQVPMKMLRWETLSIT
 WGEYEFPMNKSSTYKRYRMQFYKISFDLKKIKDLNKEIYFLLIILAKENAK
 IIFDESAAYFSQKQAFIKLIVLKSGVASFLITSETDAIKFSDEPIILKDK
 CFTVKNESLISKSSSDKVFANINCKPEKDPKPNLFEEFPMYDVSFLNKKR
 VLGITGEAVIKTWKELFLGELFVGCIDKIDIRYERINITECKAGFLPLGILNLPD
 NSSILDNFLAKFMFNENKIFIKOSTINQIKDFEKKMEFYSEKIRILYKSLAFSG
 GTLKKFALRYEMYLAKSFLICFPLSNIDHKAVMESVAIRNYSKEKPVLLITSLNDE
 LLLISDNIAMKGEVLNWSREKISKELKELFL"
 complement(7194..8246)
 /gene="BB0319"
 complement(7194..8246)
 /note="similar to GB:012861 PID:551522 percent identity:
 26.33; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="exported protein (tpn38b)"

gene
 CDS

Query Match 66.7%; Score 16; DB 1; Length 11199;
 Best Local Similarity 100.0%; Pred. No. 85;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcagattatttg 16
 |||||
 Db 5543 TCAGATTATTGCGG 5528

RESULT 15
 AC017994/c
 LOCUS
 DEFINITION AC017994 23091 bp DNA HTG 09-DEC-1999
 Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
 pieces.
 AC017994
 VERSION AC017994.1 GI:6553196
 KEYWORDS HTG: HTGS_PHASE2.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 23091)
 Adams, M. and Venter, J.C.
 Direct Submission
 Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
 Rockville, MD, USA

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 This sequence was identified as CDW:10212930 by the submitter.
 For more information on this record e-mail to fly@celera.com.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced

* By the finished sequence as soon as it is available and
 * the accession number will be preserved.
 FEATURES
 source location/Qualifiers
 1..23091
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 BASE COUNT 7147 a 4727 c 4430 g 6787 t
 ORIGIN

Query Match 66.7%; Score 16; DB 65; Length 23091;
 Best Local Similarity 100.0%; Pred. No. 82;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 agattatttg 18
 |||||
 Db 21141 AGATTATTGCGCT 21126

Search completed: November 15, 2001, 06:17:26
 Job time: 8011 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: November 15, 2001, 05:55:52 ; Search time 49.19 Seconds

(Without alignments)
82,875 Million cell updates/sec

Title: US-09-663-020-8
Perfect score: 18
Sequence: 1 ggcggagcttcagtgag 18

Scoring table: OLIGO-NUC
Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 residues

Word size : 0

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued Patents, NA:*

1: /cgn2_6/ptodata/2/lna/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/lna/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/lna/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/lna/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/lna/PCITUS.COMB.seq:*
6: /cgn2_6/ptodata/2/lna/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18	100.0	320	1	US-08-629-939-5
2	18	100.0	320	1	US-08-759-873-5
3	18	100.0	471	4	US-09-018-584A-6
4	18	100.0	631	4	US-09-385-982-354
5	18	100.0	774	3	US-08-735-587-20
6	18	100.0	863	3	US-08-943-731-171
7	18	100.0	1050	3	US-08-755-587-21
8	18	100.0	1318	2	US-08-439-814-3
9	18	100.0	1688	2	US-08-439-814-2
10	18	100.0	1712	2	US-09-058-389A-12
11	18	100.0	2090	2	US-08-439-814-1
12	18	100.0	2236	3	US-08-829-525-23
13	18	100.0	2236	4	US-08-609-583A-23
14	18	100.0	2236	4	US-08-937-399-23
15	18	100.0	2343	2	US-09-031-392-1
16	18	100.0	2343	3	US-09-299-549-1
17	18	100.0	2589	6	5212286-1
18	18	100.0	2886	6	US-08-687-080-55
19	18	100.0	3286	4	US-09-211-417-2
20	18	100.0	3647	1	US-07-914-281-7
21	18	100.0	3647	1	US-08-393-246-7
22	18	100.0	3647	1	US-08-525-058A-7
23	18	100.0	3647	1	US-08-696-731-7
24	18	100.0	3647	4	US-09-042-531-7
25	18	100.0	3647	5	PCT-US91-00899-4
26	18	100.0	3742	1	US-08-694-915-5
27	18	100.0	4668	4	US-09-045-301-1

28	18	100.0	5375	3	US-08-757-223-7	Sequence 7, Appl
29	18	100.0	5835	4	US-09-033-333-3	Sequence 3, Appl
30	18	100.0	5836	1	US-08-380-916-1	Sequence 1, Appl
31	18	100.0	5836	3	US-08-721-690-1	Sequence 1, Appl
32	18	100.0	5836	3	US-08-691-581-1	Sequence 1, Appl
33	18	100.0	5836	4	US-09-033-333-2	Sequence 2, Appl
34	18	100.0	6063	4	US-08-195-744-4	Sequence 4, Appl
35	18	100.0	6063	2	US-08-788-279-4	Sequence 4, Appl
36	18	100.0	6354	3	US-09-058-389A-5	Sequence 5, Appl
37	18	100.0	6678	3	US-08-816-617A-1	Sequence 1, Appl
38	18	100.0	8517	3	US-08-827-208-1	Sequence 1, Appl
39	18	100.0	8517	4	US-09-500-358-1	Sequence 1, Appl
40	18	100.0	8517	4	US-09-498-809-1	Sequence 1, Appl
41	18	100.0	11613	1	US-08-484-044-10	Sequence 10, Appl
42	18	100.0	11811	4	US-09-078-294-7	Sequence 7, Appl
43	18	100.0	13158	2	US-08-687-080-105	Sequence 105, App
44	18	100.0	14636	4	US-09-173-914-6	Sequence 6, Appl
45	18	100.0	14796	4	US-08-975-080-35	Sequence 35, Appl

ALIGNMENTS

RESULT 1
US-08-629-939-5/c
Sequence 5, Application US/08629939
Patent No. 5645995
GENERAL INFORMATION:
APPLICANT: Kieback, Dirk G.
TITLE OF INVENTION: METHODS FOR DIAGNOSING AN INCREASED RISK OF BREAST OR OVARIAN CANCER
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS
STREET: 2100 Pennsylvania Avenue, N.W., Suite 800
CITY: Washington, D.C.
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/629,939
FILING DATE: 12-APRIL-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: KIT, Gordon
REGISTRATION NUMBER: 30,764
REFERENCE/DOCKET NUMBER: A-6612
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 293-7860
TELEFAX: (202) 293-7060
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 320 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
HYPOTHETICAL: NO
US-08-629-939-5

Query Match 100.0%; Score 18; DB 1; Length 320;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcggagcttcagtgag 18
|||||
Db 96 ggcggagcttcagtgag 79

RESULT 2
US-08-759-873-5/C
Sequence 5, Application US/08759873
Patent No. 563885
GENERAL INFORMATION:
APPLICANT: Kieback, Dirk G.
TITLE OF INVENTION: METHODS FOR DIAGNOSING AN INCREASED RISK
TITLE OF INVENTION: OF BREAST OR OVARIAN CANCER
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: SUGHRUE, MIOM, ZINN, MACPEAK & SEAS
STREET: 2100 Pennsylvania Avenue, N.W., Suite 800
STATE: Washington, D.C.
COUNTRY: U.S.A.
ZIP: 20037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/759,873
FILING DATE: 12-APRIL-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: KIT, Gordon
REGISTRATION NUMBER: 30,764
REFERENCE/DOCKET NUMBER: A-6612
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 293-7060
TELEFAX: (202) 293-7860
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 320 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
HYPOTHETICAL: NO
US-08-759-873-5

Query Match 100.0%; Score 18; DB 1; Length 320;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggcggagcttgagtgag 18
|||||

DB 96 ggcggagcttgagtgag 79

RESULT 3
US-09-018-584A-6
Sequence 6, Application US/09018584A
Patent No. 6238863
GENERAL INFORMATION:
APPLICANT: Schumm, James W.
APPLICANT: Bachner, Jeffery W.
TITLE OF INVENTION: MATERIALS AND METHODS FOR
IDENTIFYING AND ANALYZING INTERMEDIATE TANDEM
REPEAT DNA MARKERS
TITLE OF INVENTION: REPEAT DNA MARKERS
NUMBER OF SEQUENCES: 147
CORRESPONDENCE ADDRESS:
ADDRESSEE: Promega Corporation
STREET: 2800 Woods Hollow Road
CITY: Madison
STATE: Wisconsin
COUNTRY: U.S.A.
ZIP: 53711-5399
COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette - 3.5 inch, 1.44 MB
COMPUTER: IBM compatible PC
OPERATING SYSTEM: Windows 95
SOFTWARE: Word 97 (DOS text format)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/018,584A
FILING DATE: 04-Feb-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Grady J. Frenchick
REGISTRATION NUMBER: 29,018
REFERENCE/DOCKET NUMBER: 16026,9180
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 257-3501
TELEFAX: (608) 257-2275
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 471bp
TYPE: Nucleic Acid
STRANDEDNESS: Double
TOPOLOGY: Circular
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: no
IMMEDIATE SOURCE:
LIBRARY: plasmid, pGem3Zf(+)
CLONE: C390
US-09-018-584A-6

Query Match 100.0%; Score 18; DB 4; Length 471;
Best Local Similarity 100.0%; Pred. No. 0.1;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggcggagcttgagtgag 18
|||||

DB 154 ggcggagcttgagtgag 171

RESULT 4
US-09-385-982-354/C
Sequence 354, Application US/09385982
Patent No. 6262334
GENERAL INFORMATION:
APPLICANT: ENDEGE, WILSON O., ET AL.
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CCDNA-260XX
CURRENT APPLICATION NUMBER: US/09/385,982
CURRENT FILING DATE: 1999-08-30
EARLIER APPLICATION NUMBER: 09/328,111
EARLIER FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: 60/117,393
EARLIER FILING DATE: 1999-01-27
EARLIER APPLICATION NUMBER: 60/098,639
EARLIER FILING DATE: 1998-08-31
NUMBER OF SEQ ID NOS: 544
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 354
LENGTH: 631
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)...(631)
OTHER INFORMATION: n - A,T,C or G
US-09-385-982-354

Query Match 100.0%; Score 18; DB 4; Length 631;
Best Local Similarity 100.0%; Pred. No. 0.1;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggcggagcttgagtgag 18

;; TITLE OF INVENTION: Materials and methods relating to the
;; TITLE OF INVENTION: Identification and sequencing of the BRCA2 cancer
;; TITLE OF INVENTION: Susceptibility gene and uses thereof.
;; NUMBER OF SEQUENCES: 222
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Bell Seltzer Park & Gibson
;; STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
;; CITY: Raleigh
;; STATE: NC
;; COUNTRY: USA
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/755,587
;; FILING DATE: 25-NOV-1996
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: GB 9523959.6
;; FILING DATE: 23-NOV-1995
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: GB 9525555.0
;; FILING DATE: 14-DEC-1995
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: GB 9617961.9
;; FILING DATE: 28-AUG-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Kenneth D Sibley
;; REGISTRATION NUMBER: 31,665
;; REFERENCE/DOCKET NUMBER: 5405-135
;; INFORMATION FOR SEQ ID NO: 21:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1050 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; NAME/KEY: CDS
;; LOCATION: 502..550
;; FEATURE:
;; NAME/KEY: exon
;; LOCATION: 501..550
;; US-08-755-587-21
Query Match 100.0%; Score 18; DB 3; Length 1050;
Best Local Similarity 100.0%; Pred. No. 0.096;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 ggcggagcttgcagtga 18
|||
Db 29 ggcggagcttgcagtga 12
RESULT 8
US-08-439-814-3
;; Sequence 3, Application US/08439814
;; Patent No. 5968735
;; GENERAL INFORMATION:
;; APPLICANT: STEIN, Ulrike
;; APPLICANT: WALTHER, Wolfgang
;; TITLE OF INVENTION: VECTOR FOR THE EXPRESSION OF
;; TITLE OF INVENTION: THERAPY-RELEVANT GENES
;; NUMBER OF SEQUENCES: 7
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: NIKAIKO, MARCELSTEIN, MURRAY & ORAM LLP
;; STREET: 655 Fifteenth Street, N. W., Suite 330 G
;; CITY: Washington
;; STATE: DC
;; COUNTRY: USA

;; ZIP: 20005-5701
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/439,814
;; FILING DATE: 12-MAY-1995
;; CLASSIFICATION: 514
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: DE P 4238778.7
;; FILING DATE: 12-NOV-1992
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: DE PCT/DE93/01086
;; FILING DATE: 10-NOV-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: KLESNER, Sharon N.
;; REGISTRATION NUMBER: 36,335
;; REFERENCE/DOCKET NUMBER: P1614-5015
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 202/638-5000
;; TELEFAX: 202/638-4810
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.30B (EPO)
;; APPLICATION NUMBER: DE P4238778.7
;; FILING DATE: 12-NOV-1992
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: WO PCT/DE93/01086
;; FILING DATE: 10-NOV-1993
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1318 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; US-08-439-814-3
Query Match 100.0%; Score 18; DB 2; Length 1318;
Best Local Similarity 100.0%; Pred. No. 0.094;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 ggcggagcttgcagtga 18
|||
Db 5 ggcggagcttgcagtga 22
RESULT 9
US-08-439-814-2
;; Sequence 2, Application US/08439814
;; Patent No. 5968735
;; GENERAL INFORMATION:
;; APPLICANT: STEIN, Ulrike
;; APPLICANT: WALTHER, Wolfgang
;; TITLE OF INVENTION: VECTOR FOR THE EXPRESSION OF
;; TITLE OF INVENTION: THERAPY-RELEVANT GENES
;; NUMBER OF SEQUENCES: 7
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: NIKAIKO, MARCELSTEIN, MURRAY & ORAM LLP
;; STREET: 655 Fifteenth Street, N. W., Suite 330 G
;; CITY: Washington
;; STATE: DC
;; COUNTRY: USA
;; ZIP: 20005-5701
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/439,814
FILING DATE: 12-MAY-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE P 4238778.7
FILING DATE: 12-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE PCT/DE93/01086
FILING DATE: 10-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: KLESNER, Sharon N.
REGISTRATION NUMBER: 36,335
REFERENCE/DOCKET NUMBER: P1614-5015
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202/638-5000
TELEFAX: 202/638-4810
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B (EPO)
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE P4238778.7
FILING DATE: 12-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/DE93/01086
FILING DATE: 10-NOV-1993
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1688 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-439-814-2

Query Match 100.0% Score 18; DB 2; Length 1688;
Best Local Similarity 100.0%; Pred. No. 0.092;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcggagcttcagtgag 18
|||||
DB 375 ggcggagcttcagtgag 392

RESULT 10
US-09-058-389A-12/C
Sequence 12, Application US/09058389A
Patent No. 6130065
GENERAL INFORMATION:
APPLICANT: Belt, Judith A.
APPLICANT: Crawford, Charles R.
TITLE OF INVENTION: A NITROBENZYLIMERCAPTOPURINERIBOSIDE
TITLE OF INVENTION: (NMBR)-INSENSITIVE, EQUILIBRATIVE, NUCLEOSIDE TRANSPORT
TITLE OF INVENTION: PROTEIN, NUCLEIC ACIDS ENCODING THE SAME AND METHODS OF
TITLE OF INVENTION: USE
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESSES:
ADDRESSEE: David A. Jackson, Esq.
STREET: 411 Hackensack Ave, Continental Plaza, 4th
STREET: Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/058,389A
FILING DATE: April 9, 1998
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 1340-1-013N
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-343-1684
TELEFAX: 201-487-5800
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1712 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "intron 2"
HYPOTHETICAL: NO
US-09-058-389A-12

Query Match 100.0% Score 18; DB 3; Length 1712;
Best Local Similarity 100.0%; Pred. No. 0.091;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcggagcttcagtgag 18
|||||
DB 456 ggcggagcttcagtgag 439

RESULT 11
US-08-439-814-1
Sequence 1, Application US/08439814
Patent No. 5968735
GENERAL INFORMATION:
APPLICANT: STEIN, Urike
APPLICANT: WALTHER, Wolfgang
TITLE OF INVENTION: VECTOR FOR THE EXPRESSION OF
TITLE OF INVENTION: THERAPY-RELEVANT GENES
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESSES:
ADDRESSEE: MIKAIKO, MARCELSTEIN, MURRAY & ORAM LLP
STREET: 655 Fifteenth Street, N. W., Suite 330 G
CITY: Street Lobby
STATE: Washington
COUNTRY: USA
ZIP: 20005-5701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/439,814
FILING DATE: 12-MAY-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE P 4238778.7
FILING DATE: 12-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE PCT/DE93/01086
FILING DATE: 10-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: KLESNER, Sharon N.
REGISTRATION NUMBER: 36,335
REFERENCE/DOCKET NUMBER: P1614-5015
TELECOMMUNICATION INFORMATION:

TELEPHONE: 202/638-5000
TELEFAX: 202/638-4810
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B (EFO)
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE P4238778.7
FILING DATE: 12-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/DE93/01086
FILING DATE: 10-NOV-1993
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2090 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-439-814-1

Query Match 100.0%; Score 18; DB 2; Length 2090;
Best Local Similarity 100.0%; Pred. No. 0.09;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gccgagcttcactgag 18
|||||
Db 777 gccgagcttcactgag 794

RESULT 12
US-08-829-525-23/c
Sequence 23, Application US/08829525
Patent No. 6084083
GENERAL INFORMATION:
APPLICANT: Levinson, Douglas A.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF IMMUNE DISORDERS
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036/2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/829,525
FILING DATE: 28-MAR-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/609,583
FILING DATE: 01-MAR-1996
APPLICATION NUMBER: US 08/487,748
FILING DATE: 07-JUN-1995
APPLICATION NUMBER: US 08/398,633
FILING DATE: 03-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-081
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:
LENGTH: 2236 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 42...944
OTHER INFORMATION: Human 200 gene nucleotide
US-08-829-525-23

Query Match 100.0%; Score 18; DB 3; Length 2236;
Best Local Similarity 100.0%; Pred. No. 0.089;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gccgagcttcactgag 18
|||||
Db 1472 gccgagcttcactgag 1455

RESULT 13
US-08-609-583A-23/c
Sequence 23, Application US/08609583A
Patent No. 6204371
GENERAL INFORMATION:
APPLICANT: Levinson, Douglas A.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF IMMUNE DISORDERS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036/2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/609,583A
FILING DATE: 01-MAR-1996
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/487,748
FILING DATE: 07-JUN-1995
APPLICATION NUMBER: US 08/398,633
FILING DATE: 03-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-048
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 2236 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 42...944
OTHER INFORMATION: Human 200 gene nucleotide
OTHER INFORMATION: sequence

US-08-609-583A-23

Query Match 100.0%; Score 18; DB 4; Length 2236;
Best Local Similarity 100.0%; Pred. No. 0.089;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggcgagacttcagtgag 18
|||||
Db 1472 GCGGAGCTTCAGTGAG 1455

RESULT 14

US-08-937-399-23/c
; Sequence 23, Application US/08937399
; Patent No. 6288218

GENERAL INFORMATION:

APPLICANT: Levinson, Douglas A.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE

TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF IMMUNE DISORDERS

NUMBER OF SEQUENCES: 37

CORRESPONDENCE ADDRESS:

ADDRESSER: Pennie & Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036/2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/937,399

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/609,583

FILING DATE: 01-MAR-1996

APPLICATION NUMBER: US 08/487,748

FILING DATE: 07-JUN-1995

APPLICATION NUMBER: US 08/398,633

FILING DATE: 03-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Cortez, Laura A.

REGISTRATION NUMBER: 30,742

REFERENCE/DOCKET NUMBER: 7853-048

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-790-9090

TELEFAX: 212-869-8864

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:

LENGTH: 2236 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA

FEATURE:

NAME/KEY: Coding Sequence

LOCATION: 42..944

OTHER INFORMATION: Human 200 gene nucleotide

OTHER INFORMATION: sequence

US-08-937-399-23

Db 1472 GCGGAGCTTCAGTGAG 1455

RESULT 15

US-09-031-392-1
; Sequence 1, Application US/09031392
; Patent No. 5942398

GENERAL INFORMATION:

APPLICANT: Tartaglia, Louis A.

APPLICANT: Meng, Xun

TITLE OF INVENTION: NUCLEIC ACID MOLECULES

TITLE OF INVENTION: ENCODING GLOTEX AND USES THEREOF

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSER: Fish & Richardson P.C.

STREET: 225 Franklin Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows95

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/031,392

FILING DATE: 26-FEB-1998

ATTORNEY/AGENT INFORMATION:

NAME: Melkijohn, Ph.D., Anita L.

REGISTRATION NUMBER: 35,283

REFERENCE/DOCKET NUMBER: 07334/072001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617/542-5070

TELEFAX: 617/542-8906

TELEX: 200154

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2343 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: Coding Sequence

LOCATION: 73...1761

US-09-031-392-1

Query Match 100.0%; Score 18; DB 2; Length 2343;
Best Local Similarity 100.0%; Pred. No. 0.089;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggcgagacttcagtgag 18
|||||
Db 2233 GCGGAGCTTCAGTGAG 2250Search completed: November 15, 2001, 05:55:53
Job time: 6744 secQuery Match 100.0%; Score 18; DB 4; Length 2236;
Best Local Similarity 100.0%; Pred. No. 0.089;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggcgagacttcagtgag 18
|||||
